

**IBM SPSS Analytic Server  
V3.0.1**

## **安装和配置指南**

**IBM**

**注释**

在使用本信息及其支持的产品之前，请先阅读第 51 页的『声明』中的信息。

**产品信息**

此版本是用于 IBM SPSS Analytic Server 的 V3.0.1 以及后续发行版和修订版，直至在新版本中另有说明为止。

# 目录

<b>第 1 章 先决条件</b> . . . . .	<b>1</b>	启用 HCatalog 数据源. . . . .	31
<b>第 2 章 Ambari 安装和配置</b> . . . . .	<b>3</b>	更改 Analytic Server 所使用的端口 . . . . .	32
特定于 Ambari 的先决条件 . . . . .	3	高可用性 Analytic Server . . . . .	32
在 Ambari 上安装 . . . . .	3	针对小型数据优化 JVM 选项 . . . . .	32
脱机安装. . . . .	4	迁移. . . . .	32
在由外部管理的 MySQL 环境中安装 Analytic Server . . . . .	5	在 Cloudera 上卸载 Analytic Server. . . . .	33
配置 . . . . .	6	<b>第 4 章 MapR 安装和配置.</b> . . . . .	<b>35</b>
安全性 . . . . .	6	MapR 概述 . . . . .	35
启用对 Essentials for R 的支持 . . . . .	11	在 MapR 上安装 Analytic Server. . . . .	35
启用关系数据库源 . . . . .	12	配置 MapR . . . . .	38
启用 HCatalog 数据源. . . . .	13	启用数据库回送 . . . . .	38
更改 Analytic Server 所使用的端口 . . . . .	13	启用 Apache Hive . . . . .	38
高可用性 Analytic Server . . . . .	14	启用 Apache HBase . . . . .	39
针对小型数据优化 JVM 选项 . . . . .	14	启用 Apache Spark . . . . .	40
更新客户机依赖关系 . . . . .	14	启用功能部件标志 . . . . .	40
配置 Apache Knox. . . . .	15	启用 R . . . . .	41
升级与迁移 . . . . .	18	启用 LZO . . . . .	41
卸载 . . . . .	21	为 MapR 设置 IBM SPSS Analytic Server 集群 . . . . .	42
卸载 Essentials for R . . . . .	21	卸载 MapR . . . . .	42
<b>第 3 章 Cloudera 安装和配置</b> . . . . .	<b>23</b>	在 MapR 上迁移 IBM SPSS Analytic Server . . . . .	42
Cloudera 概述 . . . . .	23	MapR 故障诊断 . . . . .	43
特定于 Cloudera 的先决条件 . . . . .	23	<b>第 5 章 配置 IBM SPSS Modeler 以用于 IBM SPSS Analytic Server</b> . . . . .	<b>45</b>
为 Analytic Server 配置 MySQL . . . . .	23	<b>第 6 章 故障诊断</b> . . . . .	<b>47</b>
在 Cloudera 上安装 . . . . .	24	<b>声明</b> . . . . .	<b>51</b>
配置 Cloudera . . . . .	25	商标 . . . . .	52
安全性 . . . . .	25		
启用对 Essentials for R 的支持 . . . . .	30		
启用关系数据库源 . . . . .	30		



---

## 第 1 章 先决条件

在安装 Analytic Server 之前，请复审以下信息。

### 系统需求

有关最新的系统需求信息，请使用 IBM 技术支持站点上的详细系统需求报告：<http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html>。在此页面上：

1. 输入 SPSS Analytic Server 作为产品名称并单击**搜索**。
2. 选择想要的版本和报告范围，然后单击**提交**。

### Power 系统

确保在集群的所有主机上的 PATH 中安装并包含 IBM XLC 和 XLF 编译器。

您可以在以下 Web 站点找到有关获取这些编译器的许可证的更多信息：

- XL C for Linux：<http://www-03.ibm.com/software/products/en/xlcpp-linux>
- XL Fortran for Linux：<http://www-03.ibm.com/software/products/en/xlfortran-linux>

### Hive/HCatalog

如果您计划使用 NoSQL 数据源，请针对远程访问配置 Hive 和 HCatalog。另外，请确保 `hive-site.xml` 包含 `hive.metastore.uris` 属性，此属性具有指向活动 Thrift Hive Metastore 服务器的格式 `thrift://<host_name>:<port>`。请参阅 Hadoop 发行版文档以获取详细信息。

### 元数据存储库

缺省情况下，Analytic Server 安装并使用 MySQL 数据库。另外，您可以配置 Analytic Server 以使用现有 DB2 安装。无论选择何种数据库类型，都必须采用 UTF-8 编码。

### MySQL

MySQL 的缺省字符集取决于版本和操作系统。使用以下步骤以确定您的 MySQL 安装是否设置为 UTF-8。

1. 确定 MySQL 的版本。

```
mysql -V
```

2. 通过从 MySQL 命令行界面运行以下查询来确定 MySQL 的缺省字符集。

```
mysql>show variables like 'char%';
```

如果字符集已设置为 UTF-8，那么无需进一步更改。

3. 通过从 MySQL 命令行界面运行以下查询来确定 MySQL 的缺省排序规则。

```
mysql>show variables like 'coll%';
```

如果排序规则已设置为 UTF-8，那么无需进一步更改。

4. 如果缺省字符集或排序规则不是 UTF-8，请参阅 MySQL 文档以获取有关如何编辑 `/etc/my.cnf` 的详细信息，然后重新启动 MySQL 守护程序以将字符集更改为 UTF-8。

**DB2** 有关配置 DB2 的更多信息，请参阅 Knowledge Center [http://www-01.ibm.com/support/knowledgecenter/SSEPGG\\_10.5.0/com.ibm.db2.luw.kc.doc/welcome.html](http://www-01.ibm.com/support/knowledgecenter/SSEPGG_10.5.0/com.ibm.db2.luw.kc.doc/welcome.html)。

### 高可用性集群

## 负载均衡器

高可用性集群应具有支持会话亲缘关系（有时候也称为粘滞会话）的负载均衡器。Analytic Server 标识包含 cookie"request-token"的会话。这标识要在应用程序控制的会话亲缘关系中使用的用户登录持续时间的会话。请参阅特定负载均衡器的文档以获取如何支持会话亲缘关系的详细信息。

---

## 第 2 章 Ambari 安装和配置

---

### 特定于 Ambari 的先决条件

除了一般的先决条件，还需查看以下信息。

**服务** Analytic Server 作为 Ambari 服务进行安装。在安装 Analytic Server 之前，您应该确保 HDFS、YARN、MapReduce2、Hive 和 Zookeeper 都作为 Ambari 进行添加。

#### 无密码 SSH

在 Analytic Metastore 主机和集群中的所有主机之间为 root 用户设置无密码 SSH。

---

### 在 Ambari 上安装

基本过程是在 Ambari 集群中的主机上安装 Analytic Server 文件，然后将 Analytic Server 添加为 Ambari 服务。以下是更详细的步骤。

1. 浏览到 [IBM Passport Advantage® Web 站点](#)，并将特定于您的堆栈、堆栈版本和硬件体系结构的自抽取二进制文件下载到 Ambari 集群中的主机。
2. 执行自解压缩二进制文件并遵循指示信息以（可选）查看许可证，接受许可证，然后选择联机或脱机安装。

#### 联机安装

如果 Ambari 服务器主机以及集群中的所有节点都能够访问 <http://ibm-open-platform.ibm.com>，那么选择联机安装。

[仅限 *GPFS (Spectrum Scale)*] 下载文件 [http://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/x86\\_64/IBM-SPSS-AnalyticServer-3.0.1.0.repo](http://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/x86_64/IBM-SPSS-AnalyticServer-3.0.1.0.repo) (x86) 或 <https://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/ppc64le/IBM-SPSS-AnalyticServer-3.0.1.0.repo> (ppc64le) 并将其移至添加 Analytic Server Metastore 作为服务的每个节点上的文件夹 `/etc/yum.repos.d` (RHEL、CentOS) 或 `/etc/zypp/repos.d` (SLES)。

#### 脱机安装

如果 Ambari 服务器主机无法访问因特网，请选择脱机安装。有关详细信息，请参阅第 4 页的『脱机安装』。

3. 重新启动 Ambari 服务器。

```
ambari-server restart
```

4. 登录 Ambari 服务器并通过 Ambari UI 将 Analytic Server 作为服务进行安装。

#### 元数据存储库

缺省情况下，Analytic Server 使用 MySQL 来跟踪数据源、项目和租户的相关信息。在安装期间，您需要提供 Analytic Server 和 MySQL 之间的 JDBC 连接中使用的用户名 (`metadata.repository.user.name`) 和密码 `metadata.repository.password`。安装程序会在 MySQL 数据库中创建用户，但该用户特定于 MySQL 数据库并且无需是现有 Linux 或 Hadoop 用户。

要将元数据存储库更改为 DB2，请执行下列步骤。

注：您无法在安装完成后更改元数据存储库。

- a. 确保 DB2 安装在另一台机器上。有关更多信息，请参阅主题第 1 页的第 1 章，『先决条件』的元数据存储库部分。

- b. 在"Ambari 服务"选项卡中, 浏览至 Analytic Server 服务的"配置"选项卡。
- c. 打开 **Advanced analytics-env** 部分。
- d. 将 **as.database.type** 的值从 `mysql` 更改为 `db2`。
- e. 打开 **Advanced analytics-meta** 部分。
- f. 将 **metadata.repository.driver** 的值从 `com.mysql.jdbc.Driver` 更改为 `com.ibm.db2.jcc.DB2Driver`。
- g. 将 **metadata.repository.url** 的值更改为 `jdbc:db2://{DB2_HOST}:{PORT}/{DBName};currentSchema={SchemaName}`; 其中
  - {DB2\_HOST} 是安装了 DB2 的服务器的主机名
  - {PORT} 是 DB2 正在侦听的端口
  - {SchemaName} 是未使用的可用模式。如果您不确定要输入什么值, 请咨询 DB2 管理员。
- h. 在 **metadata.repository.user.name** 和 **metadata.repository.password** 中提供有效的 DB2 凭证。
- i. 单击**保存**。

#### 安装后不应进行更改的配置设置

在安装后请勿更改以下设置, 否则 Analytic Server 将无法工作。

- Analytic\_Server\_User
  - Analytic\_Server\_UserID
  - as.database.type
  - metadata.repository.driver
  - distrib.fs.root
5. 现在, 您已有 Analytic Server 的正常运行示例。可以选择进行进一步配置。有关配置和管理 Analytic Server 的更多信息, 请参阅主题: 第 6 页的『配置』。有关将现有配置迁移到新安装的信息, 请参阅主题: 第 18 页的『升级与迁移』。
  6. 打开 Web 浏览器并输入地址 `http://<host>:<port>/analyticserver/admin/ibm`, 其中 `<host>` 是 Analytic Server 主机的地址, 而 `<port>` 是 Analytic Server 正在侦听的端口。缺省情况下, 此端口为 9080。此 URL 将打开 Analytic Server 控制台的登录对话框。以 Analytic Server 管理员身份登录。缺省情况下, 此用户标识是 `admin` 并且密码为 `admin`。

## 脱机安装

脱机安装将下载必需的 RPM 文件, 并且应该在可以访问 `https://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/` 的机器上运行。可执行二进制文件位于可用的 Ambari `<AS_INSTALLABLE_HOME>` 分发目录。将相应的 `<AS_INSTALLABLE_HOME>` 目录的全部内容复制到 Ambari 服务器主机。

1. 安装使您可以创建本地 Yum 存储库的工具。

```
yum install createrepo
```
2. 创建将用作 Analytic Server RPM 文件的存储库的新目录。请参阅以下示例。

```
mkdir /home/root/repos/IBM-SPSS-AnalyticServer/x86_64
```
3. 将必需的 Analytic Server RPM 文件复制到此目录。您需要的 RPM 文件取决于您的发行版、版本和体系结构, 如下所示。

#### BigInsights 4.2 (x86\_64)

`IBM-SPSS-AnalyticServer-ambari-2.1-BI-4.2-3.0.1.0-1.x86_64.rpm`

IBM-SPSS-AnalyticServer-3.0.1.0-1.x86\_64.rpm

### BigInsights 4.2 (PPC64LE)

IBM-SPSS-AnalyticServer-ambari-2.1-BI-4.2-3.0.1.0-1.ppc64le.rpm

IBM-SPSS-AnalyticServer-3.0.1.0-1.ppc64le.rpm

### HDP 2.4 (x86\_64)

IBM-SPSS-AnalyticServer-3.0.1.0-1.x86\_64.rpm

IBM-SPSS-AnalyticServer-ambari-2.1-HDP-2.4-3.0.1.0-1.x86\_64.rpm

4. 创建本地存储库定义。例如，在 `/etc/yum.repos.d/`（对于 RHEL、CentOS）或 `/etc/zypp/repos.d/`（对于 SLES）中创建名为 `IBM-SPSS-AnalyticServer-3.0.1.0.repo` 且包含以下内容的文件。

```
[IBM-SPSS-AnalyticServer]
name=IBM-SPSS-AnalyticServer
baseurl=file:///{{path to local repository}}
enabled=1
gpgcheck=0
protect=1
```

5. 创建本地 Yum 存储库。请参阅以下示例。

```
createrepo /home/root/repos/IBM-SPSS-AnalyticServer/x86_64
```

6. 在 root 用户命令窗口中，`cd` 至 `<AS_INSTALLABLE_HOME>/IBM-SPSS-AnalyticServer` 目录，然后运行 `run ./offLineInstall.sh`。脚本读取先前执行的二进制可执行安装命令的持久响应，并发出相应的平台命令（用于安装 rpm）。
7. 通过添加下列行，更新通常位于 `/var/lib/ambari-server/resources/stacks/$stackName/$stackVersion/repos/` 中的 Ambari 存储库文件 `repoinfo.xml`，以使用本地 Yum 存储库。

```
<os type="host_os">
  <repo>
    <baseurl>file:///{{path to local repository}}/</baseurl>
    <repoId>IBM-SPSS-AnalyticServer</repoId>
    <reponame>IBM-SPSS-AnalyticServer-3.0.1.0</reponame>
  </repo>
</os>
```

## 在由外部管理的 MySQL 环境中安装 Analytic Server

在由外部管理的 MySQL 环境中安装 Analytic Server 的过程与正常安装不同。

以下步骤说明了在由外部管理的 MySQL 环境中安装 Analytic Server 的过程。

1. 浏览到 [IBM Passport Advantage® Web 站点](#)，并将特定于您的堆栈、堆栈版本和硬件体系结构的自抽取二进制文件下载到 Ambari 集群中的主机。
2. 执行自解压缩二进制文件并遵循指示信息以（可选）查看许可证，接受许可证。
  - a. 选择联机选项。
  - b. 在提示时选择外部 **MySQL 数据库** 选项。
3. 将 `add_mysql_user.sh` 脚本从 `/opt/AS_Installable/IBM-SPSS-AnalyticServer` 复制到 `node/host`，即安装 MySQL（将用作 `AS_MetaStore`）的目录。例如 `/opt/AS_InstallTools`。
  - 在 MySQL `node/host` 上运行 `add_mysql_user.sh`。例如 `./add_mysql_user.sh -u as_user -p spss -d aedb`

### 注释：

- 用户名和密码必须与在 Ambari 配置屏幕上输入的 `AS_Metastore` 的数据库用户名和密码相匹配。
- 可以手动更新 `add_mysql_user.sh` 脚本，以发出命令（如有需要）。

- 针对受保护的（root 用户访问权）MySQL 数据库运行 `add_mysql_user.sh` 脚本，使用 `-r` 和 `-t` 参数传入 `dbuserid` 和 `dbuserid_password`。该脚本使用 `dbuserid` 和 `dbuserid_password` 来执行 MySQL 操作。
4. 重新启动 Ambari 服务器。
  5. 通常在 Ambari 控制台上添加 AnalyticServer 服务（输入步骤 3 中输入的数据库用户名和密码）。

注：必须在 **AS\_Configuration** 屏幕 (**Advanced analytics-meta**) 上更改 `metadata.repository.url` 设置，以指向 MySQL 数据库主机。例如，将 JDBC 设置 `mysql://{analytic_metastore_host}/aedb?createDatabaseIfNotExist=true` 更改为 `mysql://{MySQL_DB}/aedb?createDatabaseIfNotExist=true`

---

## 配置

安装后，您可以通过 Ambari UI 有选择地配置和管理 Analytic Server。

注：以下约定用于 Analytic Server 文件路径。

- `{AS_ROOT}` 指的是部署 Analytic Server 的位置；例如，`/opt/IBM/SPSS/AnalyticServer/{version}`。
- `{AS_SERVER_ROOT}` 指的是配置、日志和服务器文件的位置，例如，`/opt/IBM/SPSS/AnalyticServer/{version}/ae_wlpserver/usr/servers/aeserver`。
- `{AS_HOME}` 指的是 HDFS 上 Analytic Server 用作根文件夹的位置。

## 安全性

`security.config` 参数用于定义包含可以作为主体添加至 Analytic Server 系统的用户和组的注册表。

缺省情况下，基本注册表是使用单个用户 `admin` 及密码 `admin` 定义的。您可以通过编辑 `security.config` 或配置 Kerberos 来更改注册表。`security.config` 参数可以在 Analytic Server 服务的“配置”选项卡的高级 `analytics.cfg` 部分中找到。

注：如果您编辑 `security.config` 参数以变更注册表，那么需要将任何新用户作为主体添加至 Analytic Server 系统。请参阅《IBM SPSS Analytic Server 管理员指南》，以获取有关租户管理的详细信息。

## 更改基本注册表

通过基本注册表，您可以在 `security.config` 参数中定义用户和组的数据库。

缺省基本注册表类似于以下内容。

```
<basicRegistry id="basic" realm="ibm">
  <user name="admin" password="admin"/>
</basicRegistry>
```

变更后的基本注册表的示例如下所示。

```
<basicRegistry id="basic" realm="ibm">
  <user name="user1" password="{xor}Dz4sLG5tbGs="/>
  <user name="user2" password="Pass"/>
  <user name="user3" password="Pass"/>
  <user name="user4" password="Pass"/>
  <user name="admin" password="{xor}KzosKw="/>
  <group name="Development">
    <member name="user1"/>
    <member name="user2"/>
  </group>
  <group name="QA">
    <member name="user3"/>
    <member name="user4"/>
  </group>
</basicRegistry>
```

```

</group>
<group name="ADMIN">
  <member name="user1"/>
  <member name="admin"/>
</group>
</basicRegistry>

```

可以使用 {AS\_ROOT}/ae\_wlpserver/bin 的 securityUtility 工具对密码进行编码，从而对它们的值进行加密。

```

securityUtility encode changeit
  {xor}Pdc+MTg6Nis=

```

注：请参阅 [http://www-01.ibm.com/support/knowledgecenter/SSD28V\\_8.5.5/com.ibm.websphere.wlp.core.doc/ae/rwlp\\_command\\_securityutil.html](http://www-01.ibm.com/support/knowledgecenter/SSD28V_8.5.5/com.ibm.websphere.wlp.core.doc/ae/rwlp_command_securityutil.html) 以获取 securityUtility 工具的详细信息。

注：基本注册表用在沙箱环境中，但建议不要将其用于生产环境。

## 配置 LDAP 注册表

LDAP 注册表允许您使用外部 LDAP 服务器（例如，Active Directory 或 OpenLDAP）对用户进行认证。

**要点：**在 Ambari 中，必须将 LDAP 用户指定为 Analytic Server 管理员。

以下是针对 OpenLDAP 的 ldapRegistry 示例。

```

<ldapRegistry
  baseDN="ou=people,dc=aeldap,dc=org"
  ldapType="Custom"
  port="389"
  host="server"
  id="OpenLDAP"
  bindDN="cn=admin,dc=aeldap,dc=org"
  bindPassword="{xor}Dz4sLG5tbGs="
  searchTimeout="300000m"
  recursiveSearch="true">
  <customFilters
    id="customFilters"
    userFilter="(&(uid=%v)(objectClass=inetOrgPerson))"
    groupFilter="(&(cn=%v)(|(objectclass=organizationalUnit)))"
    groupMemberIdMap="posixGroup:memberUid"/>
</ldapRegistry>

```

以下示例提供使用 Active Directory 进行 Analytic Server 认证的方法：

```

<ldapRegistry id="Microsoft Active Directory" realm="ibm"
  host="host"
  port="389"
  baseDN="cn=users,dc=adtest,dc=mycompany,dc=com"
  bindDN="cn=administrator,cn=users,dc=adtest,dc=mycompany,dc=com"
  bindPassword="adminpassword"
  ldapType="Custom"
  <customFilters
    userFilter="(&(sAMAccountName=%v)(objectcategory=user))"
    groupFilter="(&(cn=%v)(objectcategory=group))"
    userIdMap="user:sAMAccountName"
    groupIdMap="*:cn"
    groupMemberIdMap="memberOf:member" />
</ldapRegistry>

```

注：使用 LDAP 查看器第三方工具通常有助于验证 LDAP 配置。

以下示例提供使用 Active Directory 进行 WebSphere Liberty 概要文件认证的方法：

```

<ldapRegistry id="ldap" realm="SampleLdapADRealm"
  host="ldapserverserver.mycity.mycompany.com" port="389" ignoreCase="true"
  baseDN="cn=users,dc=adtest,dc=mycity,dc=mycompany,dc=com"
  bindDN="cn=testuser,cn=users,dc=adtest,dc=mycity,dc=mycompany,dc=com"
  bindPassword="testuserpwd"
  ldapType="Microsoft Active Directory"
  sslEnabled="true"
  sslRef="LDAPSSLSettings">
  <activatedFilters
    userFilter="(&!(sAMAccountName=%v)(objectcategory=user))"
    groupFilter="(&!(cn=%v)(objectcategory=group))"
    userIdMap="user:sAMAccountName"
    groupIdMap="*:cn"
    groupMemberIdMap="memberOf:member" >
  </activatedFilters>
</ldapRegistry>

<ssl id="LDAPSSLSettings" keyStoreRef="LDAPKeyStore" trustStoreRef="LDAPTrustStore" />

<keyStore id="LDAPKeyStore" location="{server.config.dir}/LdapSSLKeyStore.jks"
  type="JKS" password="{xor}CDo9Hgw=" />

<keyStore id="LDAPTrustStore" location="{server.config.dir}/LdapSSLTrustStore.jks"
  type="JKS" password="{xor}CDo9Hgw=" />

```

#### 注释:

- 在 Analytic Server 中对 LDAP 的支持由 WebSphere Liberty 进行控制。有关更多信息，请参阅在 Liberty 中配置 LDAP 用户注册表。
- 使用 SSL 保护 LDAP 时，请遵循以下"配置从 Analytic Server 到 LDAP 的安全套接字层 (SSL) 连接"部分中的指示信息。

### 配置从 Analytic Server 到 LDAP 的安全套接字层 (SSL) 连接

1. 以 Analytic Server 用户身份登录每台 Analytic Server 机器，然后为 SSL 证书创建公共目录。

注：缺省情况下，as\_user 是 Analytic Server 用户；请在 Ambari 控制台中查看"管理"选项卡下的服务帐户。

2. 将密钥库和信任库文件复制到所有 Analytic Server 机器上的某一公共目录。另外，将 LDAP 客户机 CA 证书添加至信任库。以下是一些指示信息示例。

```

mkdir /home/as_user/security
cd /home/as_user/security
openssl s_client -connect <ldap-hostname>:636 -showcerts > client.cert
$JAVA_HOME/bin/keytool -import -file ./client.cert -alias ldapCA -keystore mytrust.jks
password : changeit

```

注：JAVA\_HOME 是用于 Analytic Server 启动的同一 JRE。

3. 可以使用 {AS\_ROOT}/ae\_wlpserver/bin 的 securityUtility 工具对密码进行编码，从而对它们的值进行加密。示例如下所示。

```

securityUtility encode changeit
{xor}PDc+MTg6Nis=

```

4. 登录 Ambari 控制台并使用正确的 SSL 配置设置更新 Analytic Server 配置设置 **ssl.keystore.config**。示例如下所示。

```

<ssl id="defaultSSLConfig" keyStoreRef="defaultKeyStore" trustStoreRef="defaultTrustStore"
  clientAuthenticationSupported="true"/>
  <keyStore id="defaultKeyStore" location="/home/as_user/security/mykey.jks" type="JKS"
    password="{xor}0zo5PiozKxYdEgwPDaWeDG1uDz4sLCg7"/>
  <keyStore id="defaultTrustStore" location="/home/as_user/security/mytrust.jks" type="JKS"
    password="{xor}PDc+MTg6Nis="/>

```

注：对密钥库文件和信任库文件使用绝对路径。

5. 使用正确的 LDAP 配置设置更新 Analytic Server 配置设置 **security.config**。例如，在 **ldapRegistry** 元素中，将 **sslEnabled** 属性设置为 true 并将 **sslRef** 属性设置为 defaultSSLConfig。

## 配置 Kerberos

Analytic Server 支持使用 Ambari 的 Kerberos。

注：IBM® SPSS® Analytic Server 与 Apache Knox 结合使用时，不支持 Kerberos 单点登录 (SSO)。

1. 在 Kerberos 用户存储库中，为您计划向其提供对 Analytic Server 的访问权的所有用户创建帐户。

注：如果 Analytic Server 安装使用基本注册表，那么它必须包含 Kerberos 用户帐户并使用 "-" 作为密码。示例如下所示。

```
<basicRegistry id="basic" realm="ibm">
  <user name="admin" password="-"/>
  <user name="user1" password="-"/>
  <user name="user2" password="-"/>
  <group name="group1">
    <member name="admin"/>
    <member name="user1"/>
    <member name="user2"/>
  </group>
  <group name="group2">
    <member name="admin"/>
    <member name="user1"/>
  </group>
</basicRegistry>
```

2. 在每个 Analytic Server 节点和 Hadoop 节点上，为前一步骤中创建的每个用户创建一个操作系统用户帐户。
  - 确保这些用户的 UID 在所有机器上都匹配。您可以使用 `kinit` 命令登录每个帐户来测试此内容。
  - 确保 UID 符合"用于提交作业的最小用户标识"Yarn 设置。这是 `container-executor.cfg` 中的 **min.user.id** 参数。例如，如果 **min.user.id** 为 1000，那么创建的每个用户帐户都必须具有大于或等于 1000 的 UID。
3. 在 Analytic Server 中，为所有主体在 HDFS 上创建用户主文件夹。例如，如果您将 `testuser1` 添加至 Analytic Server 系统，请在 HDFS 上创建类似于 `/user/testuser1` 的主文件夹，并确保 `testuser1` 具有对此文件夹的读写许可权。
4. [可选] 如果您计划使用 HCatalog 数据源并且 Analytic Server 与 Hive metastore 安装在不同的机器上，那么必须在 HDFS 上模拟 Hive 客户机。
  - a. 在 Ambari 控制台中，浏览至 HDFS 服务的"配置"选项卡。
  - b. 编辑 **hadoop.proxyuser.hive.groups** 参数以使其具有值 `*`，或者编辑包含所有能够登录 Analytic Server 的用户的组。
  - c. 编辑 **hadoop.proxyuser.hive.hosts** 参数以使其具有值 `*`，或者编辑 Hive metastore 和每个 Analytic Server 实例都作为服务安装在其上主机的列表。
  - d. 重新启动 HDFS 服务。

执行这些步骤并且安装了 Analytic Server 步骤后，Analytic Server 将以静默方式自动配置 Kerberos。

## 使用 Kerberos 针对单点登录 (SSO) 配置 HAProxy

1. 按照 HAProxy 文档指南配置并启动 HAProxy：<http://www.haproxy.org/#docs>
2. 针对 HAProxy 主机创建 Kerberos 原则 (`HTTP/<proxyHostname>@<realm>`) 和密钥表文件，其中，`<proxyHostname>` 是 HAProxy 主机的完整名称，而 `<realm>` 是 Kerberos 域。

3. 将密钥表文件作为 `/etc/security/keytabs/spnego_proxy.service.keytab` 复制到每个 Analytic Server 主机
4. 在每个 Analytic Server 主机上更新此文件的许可权。 示例如下所示。  

```
chown root:hadoop /etc/security/keytabs/spnego_proxy.service.keytab
chmod 440 /etc/security/keytabs/spnego_proxy.service.keytab
```
5. 打开 Amabri 控制台并更新 Analytic Server"Custom analytics.cfg"部分中的以下属性。  

```
web.authentication.kerberos.keytab=/etc/security/keytabs/spnego_proxy.service.keytab
web.authentication.kerberos.principal=HTTP/<proxy machine full name>@<realm>
```
6. 从 Amabri 控制台保存配置并重新启动所有 Analytic Server 服务。

现在，用户能够使用 Kerberos SSO 登录到 Analytic Server。

## 禁用 Kerberos

1. 在 Ambari 控制台中禁用 Kerberos。
2. 停止 Analytic Server 服务。
3. 从定制 `analytics.cfg` 中除去下列参数。

```
default.security.provider
hdfs.keytab
hdfs.user
java.security.krb5.conf
jdbc.db.connect.method.kerberos
web.authentication.kerberos.keytab
web.authentication.kerberos.principal
```

4. 单击保存，然后重新启动 Analytic Server 服务。

## 启用与 Analytic Server 控制台的安全套接字层 (SSL) 连接

缺省情况下，Analytic Server 会生成自签名证书以启用安全套接字层 (SSL)，因此您可以接受自签名证书，以通过安全端口来访问 Analytic Server 控制台。 要使 HTTPS 访问更安全，您需要安装第三方供应商证书。

要安装第三方供应商证书，请执行下列步骤。

1. 将第三方供应商密钥库和信任库复制到所有 Analytic Server 节点中的同一目录；例如，`/home/as_user/security`。

**注：**Analytic Server 用户必须具有对此目录的读访问权。

2. 在"Ambari 服务"选项卡中，浏览至 Analytic Server 服务的"配置"选项卡。
3. 编辑 `ssl.keystore.config` 参数。

```
<ssl id="defaultSSLConfig"
  keyStoreRef="defaultKeyStore"
  trustStoreRef="defaultTrustStore"
  clientAuthenticationSupported="true"/>
<keyStore id="defaultKeyStore"
  location="<KEYSTORE-LOCATION>"
  type="<TYPE>"
  password="<PASSWORD>"/>
<keyStore id="defaultTrustStore"
  location="<TRUSTSTORE-LOCATION>"
  type="<TYPE>"
  password="<PASSWORD>"/>
```

将

- `<KEYSTORE-LOCATION>` 替换为密钥库的绝对位置；例如，`/home/as_user/security/mykey.jks`

- <TRUSTSTORE-LOCATION> 替换信任库的绝对位置；例如：/home/as\_user/security/mytrust.jks
- <TYPE> 替换为证书的类型；例如：JKS、PKCS12 等。
- <PASSWORD> 替换为使用 Base64 加密格式的加密密码。对于编码，您可以使用 securityUtility；例如：/opt/ibm/spss/analyticserver/3.0/ae\_wlpserver/bin/securityUtility encode <password>

如果要生成自签名证书，您可以使用 securityUtility；例如：/opt/ibm/spss/analyticserver/3.0/ae\_wlpserver/bin/securityUtility createSSLCertificate --server=myserver --password=myspassword --validity=365 --subject=CN=mycompany,O=myOrg,C=myCountry。

有关 securityUtility 和其他 SSL 设置的更多信息，请参阅 WebSphere Liberty 概要文件文档。

4. 单击保存，然后重新启动 Analytic Server 服务。

## 启用对 Essentials for R 的支持

Analytic Server 支持对 R 模型进行评分，并支持运行 R 脚本。

为了配置对 R 的支持，成功安装 Analytic Server 后，请执行以下步骤：

1. 下载用于 IBM SPSS Modeler Essentials for R RPM 的自解压缩归档 (BIN)。Essentials for R 可供下载 (<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>)。选择特定于堆栈、堆栈版本及硬件体系结构的文件。
2. 执行自解压缩二进制文件并遵循指示信息以（可选）查看许可证，接受许可证，然后选择联机或脱机安装。

### 联机安装

如果 Ambari 服务器主机以及集群中的所有节点都能够访问 <http://ibm-open-platform.ibm.com>，那么选择联机安装。

[仅限 *GPFS (Spectrum Scale)*] 下载文件 [http://ibm-open-platform.ibm.com/repos/IBM-SPSS-ModelerEssentialsR/3.0.1.0/x86\\_64/IBM-SPSS-AnalyticServer-3.0.1.0.repo](http://ibm-open-platform.ibm.com/repos/IBM-SPSS-ModelerEssentialsR/3.0.1.0/x86_64/IBM-SPSS-AnalyticServer-3.0.1.0.repo) (x86) 或 [http://ibm-open-platform.ibm.com/repos/IBM-SPSS-ModelerEssentialsR/3.0.1.0/x86\\_64/IBM-SPSS-AnalyticServer-3.0.1.0.repo](http://ibm-open-platform.ibm.com/repos/IBM-SPSS-ModelerEssentialsR/3.0.1.0/x86_64/IBM-SPSS-AnalyticServer-3.0.1.0.repo) (ppc64le) 并将其移至添加 Analytic Server Metastore 作为服务的每个节点上的文件夹 /etc/yum.repos.d (RHEL、CentOS) 或 /etc/zypp/repos.d (SLES)。

### 脱机安装

如果 Ambari 服务器主机无法访问因特网，请选择脱机安装。脱机安装将下载必需的 RPM 文件，并且应该在可以访问 <http://ibm-open-platform.ibm.com> 的机器上运行。然后，可以将 RPM 文件复制到 Ambari 服务器主机。

- a. 将必需的 Essentials for R RPM 文件复制到 Ambari 服务器主机上的任何位置。您需要的 RPM 文件取决于您的发行版、版本和体系结构，如下所示。

#### BigInsights 4.2 (x86\_64)

IBM-SPSS-ModelerEssentialsR-ambari-2.1-BI-4.2-8.4.0.0-1.x86\_64.rpm

#### BigInsights 4.2 (PPC64LE)

IBM-SPSS-ModelerEssentialsR-ambari-2.1-BI-4.2-8.4.0.0-1.ppc64le.rpm

#### HDP 2.4 (x86\_64)

IBM-SPSS-ModelerEssentialsR-ambari-2.1-HDP-2.4-8.4.0.0-1.x86\_64.rpm

- b. 安装 RPM。例如，以下命令为在 BigInsights 4.1 上安装 Essentials for R。

```
rpm -i IBM-SPSS-ModelerEssentialsR-ambari-2.1-BI-4.2-8.4.0.0-1.x86_64.rpm
```

3. 重新启动 Ambari 服务器。

ambari-server restart

4. 登录 Ambari 服务器并通过 Ambari 控制台将 SPSS Essentials for R 作为服务进行安装。应该在安装了 Analytic Server 和 Analytic Metastore 的每个主机上安装 SPSS Essentials for R。

注：Ambari 将尝试在安装 R 之前安装 gcc-c++ 和 gcc-gfortran (RHEL) 以及 gcc-fortran (SUSE)。这些软件包声明为对 R 的 Ambari 服务定义的从属项。请确保要安装和执行 R 的服务器配置为下载 gcc-c++ 和 gcc-[g]fortran RPM，或者已安装 GCC 和 FORTRAN 编译器。如果 Essentials for R 安装失败，请在安装 Essentials for R 之前手动安装这些包。

5. 刷新 Analytic Server 服务。
6. 使用第 14 页的『更新客户机依赖关系』中的指示信息运行 update\_clientdeps 脚本。
7. 还必须在托管 SPSS Modeler 服务器的机器上安装 Essentials for R。请参阅 SPSS Modeler 文档以获取详细信息。

## 启用关系数据库源

如果您在每个 Analytic Server 主机的共享目录中提供 JDBC 驱动程序，那么 Analytic Server 可以使用关系数据库源。缺省情况下，此目录为 /usr/share/jdbc。

要更改共享目录，请执行下列步骤。

1. 在"Ambari 服务"选项卡中，浏览至 Analytic Server 服务的"配置"选项卡。
2. 打开 **Advanced analytics.cfg** 部分。
3. 在 **jdbc.drivers.location** 中，指定 JDBC 驱动程序的共享目录的路径。
4. 单击保存。
5. 停止 Analytic Server 服务。
6. 单击刷新。
7. 启动 Analytic Server 服务。

表 1. 受支持的数据库

数据库	受支持的版本	JDBC 驱动程序 JAR	供应商
Amazon Redshift	8.0.2 或更高版本	RedshiftJDBC41-1.1.6.1006.jar 或更高版本	Amazon
DashDB	Bluemix 服务	db2jcc.jar	IBM
DB2 for Linux、UNIX 和 Windows	10.5、10.1 和 9.7	db2jcc.jar	IBM
DB2 z/OS	11 和 10	db2jcc.jar 和 db2_license_cisuz.jar	IBM
Greenplum	5 和 4.2.x	postgresql.jar	Greenplum
Hive	1.1、1.2	hive-jdbc-*.jar	Apache
Netezza	7 和 6.x	nzjdbc.jar	IBM
Oracle	12c 和 11g R2 (11.2)	ojdbc6.jar 和 orai18n.jar	Oracle
SQL Server	2014、2012 和 2008 R2	sqljdbc4.jar	Microsoft
Sybase IQ	16.x、15.4 和 15.2	jconnect70.jar	Sybase
Teradata	14、14.1 和 15	tdgssconfig.jar 和 terajdbc4.jar	Teradata

注：如果在安装 Analytic Server 之前创建了 Redshift 数据源，那么需要执行以下步骤，才能使用该 Redshift 数据源。

1. 在 Analytic Server 控制台中，打开该 Redshift 数据源。
2. 选择 Redshift 数据库数据源。
3. 输入 Redshift 服务器地址。
4. 输入数据库名称和用户名。密码应自动填充。
5. 选择数据库表。

## 启用 HCatalog 数据源

Analytic Server 通过 Hive/HCatalog 提供对许多数据源的支持。一些源需要手动配置步骤。

1. 收集必需的 JAR 文件以启用数据源。请参阅下列各节以获取详细信息。
2. 将这些 JAR 文件添加至 {HIVE\_HOME}/auxlib 目录及每个 Analytic Server 节点的 /usr/share/hive 目录。
3. 重新启动 Hive Metastore 服务。
4. 刷新 Analytic Metastore 服务。
5. 重新启动 Analytic Server 服务的每个实例。

## NoSQL 数据库

Analytic Server 支持任何 NoSQL 数据库，可以从供应商处获取用于此数据库的 Hive 存储处理程序。

不需要任何额外步骤即可启用对 Apache HBase 及 Apache Accumulo 的支持。

对于其他 NoSQL 数据库，请联系联系人供应商并获取存储处理程序及相关 JAR。

## 基于文件的 Hive 表

Analytic Server 支持任何基于文件的 Hive 表，有一个内置或定制 Hive SerDe（序列化器-反序列化器）可用于此表。

用于处理 XML 文件的 Hive XML SerDe 位于 Maven Central Repository 中，网址为 <http://search.maven.org/#search%7Cga%7C1%7Chivexmlserde>。

## Apache Spark

如果要使用 Spark (V1.5 或更高版本) 和 HCatalog 输入数据源，必须在定制的 analytics.cfg 文件中手动添加 spark.version 属性。

1. 打开 Amabri 控制台，并在 Analytic Server 的定制 **analytics.cfg** 部分中添加以下属性。
  - **Key:** spark.version
  - **Value:** 输入相应的 Spark 版本号（例如 1.5）。
2. 从 Amabri 控制台保存配置并重新启动所有 Analytic Server 服务。

## 更改 Analytic Server 所使用的端口

缺省情况下，Analytic Server 对 HTTP 和 HTTPS 分别使用 9080 和 9443 端口。要更改端口设置，请执行下列步骤。

1. 在"Ambari 服务"选项卡中，浏览至 Analytic Server 服务的"配置"选项卡。

2. 打开 **Advanced analytics.cfg** 部分。
3. 在 **http.port** 和 **https.port** 中分别指定所需的 HTTP 和 HTTPS 端口。
4. 单击**保存**。
5. 重新启动 Analytic Server 服务。

## 高可用性 Analytic Server

通过将 Analytic Server 作为服务添加至集群中的多个节点，您可以实现它的高可用性。

1. 在 Ambari 控制台中，浏览至"主机"选项卡。
2. 选择尚未将 Analytic Server 作为服务运行的主机。
3. 在"摘要"选项卡上，单击**添加**并选择 Analytic Server。
4. 单击**确认添加**

## 针对小型数据优化 JVM 选项

您可以编辑 JVM 属性，以便在运行小型 (M3R) 作业时优化系统。

在 Ambari 控制台中，查看 Analytic Server 服务中"配置"选项卡的 Advanced analytics-jvm-options 部分。修改下列参数会设置托管 Analytic Server 的服务器上运行的作业的堆大小；即，并非 Hadoop。运行较小的 (M3R) 作业时，这一点很重要，并且您可能需要对这些值进行试验以优化系统。

```
-Xms512M  
-Xmx2048M
```

## 更新客户机依赖关系

本节描述如何使用 update\_clientdeps 脚本更新 Analytic Server 服务的依赖关系。

1. 以 root 用户身份登录到 Ambari 服务器主机。
2. 将目录更改为 /var/lib/ambari-server/resources/stacks/<stack-name>/<stack-version>/services/ANALYTICSERVER/package/scripts；请参阅以下示例。

```
cd "/var/lib/ambari-server/resources/stacks/HDP/2.4/services/ANALYTICSERVER/package/scripts"
```

3. 使用以下参数运行 update\_clientdeps 脚本。

**-u <ambari-user>**

Ambari 帐户用户名

**-p <ambari-password>**

Ambari 帐户用户的密码。

**-h <ambari-host>**

Ambari 服务器的主机名。

**-x <ambari-port>**

Ambari 侦听的端口。

请参阅以下示例。

```
./update_clientdeps.sh -u admin -p admin -h host.domain -x 8080
```

4. 使用以下命令重新启动 Ambari 服务器。

```
ambari-server restart
```

## 配置 Apache Knox

Apache Knox Gateway 是针对 Apache Hadoop 服务提供单点安全访问的系统。系统简化用户（访问集群数据和运行作业）和操作员（控制访问权和管理集群）的 Hadoop 安全性。Gateway 作为服务器（或服务器集群）运行，为一个或多个 Hadoop 集群提供服务。

注：IBM SPSS Analytic Server 与 Kerberos 单点登录 (SSO) 结合使用时，不支持 Apache Knox。

Apache Knox Gateway 有效地隐藏 Hadoop 集群拓扑详细信息并与 Enterprise LDAP 和 Kerberos 相集成。以下部分提供有关必需的 Apache Knox 和 Analytic Server 配置任务的信息。

要点：Analytic Server 无法安装在与 Knox 服务器相同的集群节点上。

### 先决条件

- Analytic Server 无法安装在与 Knox 服务器相同的集群节点上。
- Analytic Server 节点必须以无密码的 SSH 连接方式连接到 Knox 服务器。无密码的 SSH 连接从 Analytic Server 移动到 Knox (**Analytic Server > Knox**)。
- 要安装 Analytic Server，必须首先安装 Knox 服务。

在某些情况下，会发生意外而导致无法自动复制配置文件。在这些情况下，您必须手动复制以下配置文件：

- com.ibm.spss.knox\_0.7-3.0.0.0.jar：必须从以下 Analytic Server 位置复制该文件：

```
<Analytic_Server_Installation_Path>/ae_wlpserver/usr/servers/aeserver/apps/AE_BOOT.war/WEB-INF/lib
```

到 Knox 服务器节点：

```
/KnoxServicePath/ext
```

例如：/usr/iop/4.1.0.0/knox/ext

- rewrite.xml 和 service.xml：必须从以下 Analytic Server 位置复制这些文件：

```
<Analytic_Server_Installation_Path>/ae_wlpserver/usr/servers/aeserver/configuration/knox
```

到 Knox 服务器节点：

```
/KnoxServicePath/data/services
```

例如：/usr/iop/4.1.0.0/knox/data/services

## 配置 Ambari

必须在 Ambari 用户界面中配置 Analytic Server 服务：

1. 在 Ambari 用户界面中，浏览至 **Knox > 配置 > 高级拓扑**。当前 Knox 配置设置在内容窗口中显示。
2. 向 Knox 配置添加以下 <service>：

```
<service>  
  <role>ANALYTICSERVER</role>  
  <url>http://{analyticserver-host}:{analyticserver-port}/analyticserver</url>  
</service>
```

{analyticserver-host} 和 {analyticserver-port} 必须替换为相应的 Analytic Server 服务器名称和端口号：

- 可以在 Ambari 用户界面 (**SPSS Analytic Server > 摘要 > Analytic Server**) 中找到 {analyticserver-host} URL。
- 可以在 Ambari 用户界面 (**SPSS Analytic Server > 配置 > Advanced analytics.cfg > http.port**) 中找到 {analyticserver-port} 号。

注：在将 Analytic Server 部署到多个节点并且使用 LoadBalancer时，{analyticserver-host} 和 {analyticserver-port} 必须对应于 LoadBalancer URL 和端口号。

### 3. 重新启动 Knox 服务。

在使用 LDAP 时，Knox 缺省为提供的"演示"LDAP。您可以更改为企业 LDAP 服务器（例如，Microsoft LDAP 或 OpenLDAP）。

## 配置 Analytic Server

要将 LDAP 用于 Analytic Server，必须将 Analytic Server 配置为使用 Apache Knox 所使用的相同 LDAP 服务器。必须更新以下 Ambari 设置的 <value> 条目以反映相应的 Knox LDAP 服务器设置：

- main.ldapRealm.userDnTemplate
- main.ldapRealm.contextFactory.url

Ambari 用户界面中提供的值位于：**Knox > 配置 > 高级拓扑**。例如：

```
<param>
  <name>main.ldapRealm.userDnTemplate</name>
  <value>uid={0},ou=people,dc=hadoop,dc=apache,dc=org</value>
</param>
<param>
  <name>main.ldapRealm.contextFactory.url</name>
  <value>ldap://{knox_host_name}:33389</value>
</param>
```

更新 Knox LDAP 设置后，重新启动 Knox 服务。

**要点：**Analytic Server 管理员密码必须与 Knox 管理员密码相同。

## 配置 Apache Knox

1. 在 Knox 服务器上，创建子目录 <knox\_server>/data/service/analyticserver/3.0，然后将 service.xml 和 rewrite.xml 文件上载到此新目录。这两个文件在 Analytic Server 上，位于 <analytic\_server>/configuration/knox/analyticserver/3.0.1（例如 /opt/ibm/spss/analyticserver/3.0/ae\_wlpserver/usr/servers/aeserver/configuration/knox/analyticserver/3.0/\*.xml）
2. 在 <knox\_server>/bin 中，运行脚本 ./knoxcli.sh redeploy--cluster default
3. 将 com.ibm.spss.knoxservice\_0.7.0-\*.jar 文件上载到 <knox\_server>/ext。该文件在 Analytic Server 上，位于 <analytic\_server>/apps/AE\_BOOT.war/WEB-INF/lib/com.ibm.spss.knox\_0.7-3.0.1.0.jar（例如 /opt/ibm/spss/analyticserver/3.0/ae\_wlpserver/usr/servers/aeserver/apps/AE\_BOOT.war/WEB-INF/lib/com.ibm.spss.knox\_0.7-3.0.1.0.jar）。
4. 在 Ambari 用户界面中，浏览至 **Knox > 配置 > 高级拓扑**，以添加下列元素。

```
<service>
  <role>ANALYTICSERVER</role>
  <url>http://{AS-Host}:{AS-port}/analyticserver</url>
</service>
```

5. 在 Ambari 用户界面中，浏览至 **Knox > 配置 > 高级用户-ldif**（例如 admin、qauser1、qauser2），以添加或更新用户。
6. 浏览至 **Knox > 服务操作 > 启动演示 LDAP** 以重新启动 LDAP。
7. 重新启动 Knox 服务。

## 在 Hortonworks 数据平台 (HDP) 上安装 Apache Knox

下列步骤概述了在 HDP 集群中安装 Apache Knox 的过程。

1. 验证 Knox 用户是否存在于 HDP 集群。如果不存在，必须创建一个 Knox 用户。
2. 下载 Apache Knox 并将其解压缩到 /home/knox 目录下的文件夹中。
3. 在 HDP 中，切换到 Knox 用户，然后转至 knox 文件夹。Knox 用户必须对所有 knox 子文件夹具有 permission(RWX) 许可权。
4. 为 Analytic Server 配置 Apache Knox。有关更多信息，请参阅配置 **Apache Knox** 部分。
  - a. 在 {knox}/data/services 目录下创建 analyticserver/3.0.1 文件夹层次结构。
  - b. 从以下 Analytic Server 位置复制 rewrite.xml 和 service.xml 文件：

```
/opt/ibm/spss/analyticserver/3.0.1/ae_wlpserver/usr/servers/aeserver/configuration/knox/  
analyticserver/3.0.1
```

到 Knox 服务器节点：

```
{knox}/data/services/analyticserver/3.0.1
```

- c. 从 Analytic Server 主机复制 Knox \*.jar 文件：

```
/opt/ibm/spss/analyticserver/3.0.1/ae_wlpserver/usr/servers/aeserver/apps/AE_BOOT.war/WEB-  
INF/lib/com.ibm.spss.knox_0.7-*.jar
```

到 Knox ext 目录：

```
{knox}/ext
```

- d. 在 {knox}/conf/topologies 目录下更新 default.xml 文件，以符合以下示例：

注：如果不存在，必须创建此文件。

```
<topology>  
  <gateway>  
    <provider>  
      <role>authentication</role>  
      <name>ShiroProvider</name>  
      <enabled>true</enabled>  
      <param>  
        <name>sessionTimeout</name>  
        <value>30</value>  
      </param>  
      <param>  
        <name>main.ldapRealm</name>  
        <value>org.apache.hadoop.gateway.shirorealm.KnoxLdapRealm</value>  
      </param>  
      <param>  
        <name>main.ldapRealm.userDnTemplate</name>  
        <value>uid={0},ou=people,dc=hadoop,dc=apache,dc=org</value>  
      </param>  
      <param>  
        <name>main.ldapRealm.contextFactory.url</name>  
        <value>ldap://localhost:33389</value>  
      </param>  
      <param>  
        <name>main.ldapRealm.contextFactory.authenticationMechanism</name>  
        <value>simple</value>  
      </param>  
      <param>  
        <name>urls./**</name>  
        <value>authcBasic</value>  
      </param>  
    </provider>  
    <provider>  
      <role>identity-assertion</role>  
      <name>Default</name>  
      <enabled>true</enabled>  
    </provider>  
    <provider>  
      <role>authorization</role>
```

```

        <name>AclsAuth</name>
        <enabled>true</enabled>
    </provider>

</gateway>

<!--other service-->
<service>
    <role>ANALYTICSERVER</role>
    <!--replace the {AS-host}nas {AS-port} with real value-->
    <url>http://{AS-host}:{AS-port}/analyticserver</url>
</service>

</topology>

```

5. 运行 `{knox}/bin/knoxcli.sh`。

6. 运行 `{knox}/bin/ldap.sh start`。

注：此脚本使用端口 33389。请确保该端口当前不在使用中。

7. 运行 `{knox}/bin/gateway.sh start`。

注：此脚本使用端口 8443。请确保该端口当前不在使用中。

8. 验证安装。

a. 在 Knox URL 上针对 Analytic Server 运行 `curl` 命令：

```
curl -ikvu {username}:{password} https://{knox-host}:8443/gateway/default/analyticserver/admin
```

## 故障诊断

问题：安装后，Analytic Server 无法在 Knox 中运行。

解决方案：停止 Knox，除去 `{knox}/data/deployments/*` 目录下的所有文件，然后重新启动 Knox。

问题：无法通过 Knox 登录 Analytic Server。

解决方案：验证 `{knox}/conf/users.ldif` 中的用户。更新现有用户，或添加新的 Analytic Server 用户。Knox 用户主体和凭证必须与 Analytic Server 用户匹配。

## 启用 Apache Knox 的 Analytic Server 的 URL 结构

启用 Knox 的 Analytic Server 用户界面 URL 为 `https://{knox-host}:{knox-port}/gateway/default/analyticserver/admin`

- https 协议 - 用户必须接受证书才能在 Web 浏览器中继续。
- `knox-host` 是 Knox 主机。
- `knox-port` 是 Knox 端口号。
- URI 为 `gateway/default/analyticserver`。

---

## 升级与迁移

通过 Analytic Server，您可以进行升级或将数据和配置设置从现有 Analytic Server 迁移到新安装。

### 从 V3.0 升级到 V3.0.1

如果您已安装 Analytic Server 3.0，那么可以在此基础上将安装升级到 V3.0.1。

1. 在 Ambari 控制台中，停止 Analytic Server 服务。
2. 根据您的安装类型，请执行以下步骤。

#### 联机升级

- a. 确保 Ambari 服务器主机以及集群中的所有节点都能够访问 <http://ibm-open-platform.ibm.com>。
- b. 在每个 Analytic Server 主机上，从 [http://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/x86\\_64/IBM-SPSS-AnalyticServer-3.0.1.0.repo](http://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/x86_64/IBM-SPSS-AnalyticServer-3.0.1.0.repo) (x86) 或 <http://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/ppc64le/IBM-SPSS-AnalyticServer-3.0.1.0.repo> (ppc64le) 中下载 IBM-SPSS-AnalyticServer-3.0.1.0.repo 文件，然后将其移动到文件夹 /etc/yum.repos.d (RHEL 或 CentOS) 或者 /etc/zypp/repos.d (SLES)。

### 脱机升级

- a. 脱机升级将下载必需的 RPM 文件，并且应该在可以访问 <http://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/> 的机器上运行。
- b. 创建将用作 Analytic Server RPM 文件的存储库的新目录。请参阅以下示例：  

```
mkdir /home/root/repos/IBM-SPSS-AnalyticServer/3.0.1.0/x86_64
```
- c. 将必需的 Analytic Server RPM 文件复制到此目录。您所需的 RPM 文件取决于分发版、版本和体系结构。对于 BigInsights 4.1，以下显示的为必需文件。

表 2. BigInsights 4.2 RPM

BigInsights 4.2 (x86_64)	BigInsights 4.2 (PPC64LE)	HDP 2.4 (x86_64)
IBM-SPSS-AnalyticServer-ambari-2.1-BI-4.2-3.0.1.0-1.x86_64.rpm	IBM-SPSS-AnalyticServer-ambari-2.1-BI-4.2-3.0.1.0-1.ppc64le.rpm	IBM-SPSS-AnalyticServer-ambari-2.1-HDP-2.4-3.0.1.0-1.x86_64.rpm
IBM-SPSS-AnalyticServer-3.0.1.0-1.x86_64.rpm	IBM-SPSS-AnalyticServer-3.0.1.0-1.ppc64le.rpm	IBM-SPSS-AnalyticServer-3.0.1.0-1.x86_64.rpm

- d. 创建本地存储库定义。例如，在 /etc/yum.repos.d/ (对于 RHEL、CentOS) 或 /etc/zypp/repos.d/ (对于 SLES) 中创建名为 analyticserver.repo 的文件。

```
[IBM-SPSS-AnalyticServer]
name=IBM-SPSS-AnalyticServer-3.0.1.0
baseurl=file:///path to local repository
enabled=1
gpgcheck=0
protect=1
```

- e. 创建本地 Yum 存储库。请参阅以下示例：

```
createrepo /home/root/repos/IBM-SPSS-AnalyticServer/3.0.1.0/x86_64
```

3. 清除本地高速缓存中的 Ambari 元数据。例如，要清除 RHEL 或 CentOS 上的高速缓存，请运行以下命令：

```
sudo yum clean all
```

注：当存在两个列出的 Analytic Server 存储库时，yum 命令不生效。因此，必须重命名或删除与 Analytic Server 相关的原始 \*.repo 文件。在 SLES 上，命令将按以下所示内容进行更改：

```
sudo zypper refresh
```

4. 在每个 Analytic Server 主机上，升级 RPM。例如，要在 RHEL 或 CentOS 上升级，请运行以下命令：

```
chown -R as_user:hadoop /opt/ibm/spss/analyticserver/3.0
sudo yum upgrade IBM-SPSS-AnalyticServer
```

在 SLES 上，命令将按以下所示内容进行更改：

```
sudo zypper up IBM-SPSS-AnalyticServer
```

5. 刷新堆栈。

## BigInsights

- a. 在 Ambari 控制台中，启动 Analytic Server 服务，然后停止此服务。
- b. 执行定制的刷新操作。

## Hortonworks

导航到其中一个 Analytic Server 节点并运行以下命令：

```
sudo -u as_user /opt/ibm/spss/analyticserver/3.0/bin/refresh.sh
```

6. 仅限脱机安装。要更新通常位于 `/var/lib/ambari-server/resources/stacks/$stackName/$stackVersion/repos/` 中的 Ambari 存储库文件 `repoinfo.xml` 以使用本地 Yum 存储库，请添加下列行：

```
<os type="host_os">
  <repo>
    <baseurl>file:///{{path to local repository}}/</baseurl>
    <repoid>IBM-SPSS-AnalyticServer</repoid>
    <reponame>IBM-SPSS-AnalyticServer-3.0.1.0</reponame>
  </repo>
</os>
```

7. 清除 Zookeeper 状态。在 Zookeeper bin 目录（例如 `/usr/iop/current/zookeeper-server/bin`）中运行以下命令：

```
./zkCli.sh rmr /AnalyticServer
```

8. 在 Ambari 控制台中，启动 Analytic Server 服务。

## 迁移到新版本的 Analytic Server

如果您已安装 Analytic Server 2.0 或 2.1 并已购买 3.0，那么可以将 2.0/2.1 配置设置迁移到 3.0/3.0.1 安装。

### Restrictions:

- 如果您已安装低于 2.0 的版本，那么必须先从较低版本迁移到 V2.0/V2.1，然后再从 V2.0/V2.1 迁移到 V3.0/V3.0.1。
- 2.0/2.1 和 3.0/3.0.1 安装无法在同一 Hadoop 集群中共存。如果您将 3.0/3.0.1 安装配置为使用 2.0/2.1 安装所用的 Hadoop 集群，那么 2.0/2.1 安装将不再起作用。

## 迁移步骤，2.0/2.1 到 3.0/3.0.1

1. 根据第 3 页的『在 Ambari 上安装』中的指示信息，安装新的 Analytic Server。
2. 将分析根目录从旧安装复制到新安装。

- a. 如果您不确定分析根目录的位置，请运行 `hadoop -fs ls`。分析根目录的路径的格式为 `/user/aeuser/analytic-root`，其中 `aeuser` 是拥有分析根目录的用户标识。

- b. 将分析根目录的所有权从 `aeuser` 更改为 `as_user`

```
hadoop dfs -chown -R {as_user:{group}} {path to 2.0/2.1 analytic-root}
```

注：如果您打算在迁移后使用现有的 Analytic Server 安装，请在 HDFS 生成 `analytic-root` 目录的副本，然后更改该目录副本的所有权。

- c. 以 `as_user` 身份登录新 Analytic Server 安装版本的主机。如果 `/user/as_user/analytic-root` 目录存在，请将其删除。

- d. 运行下列复制脚本。

```
hadoop distcp hftp://{{host of 2.0/2.1 namenode}}:50070/{{path to 2.0/2.1 analytic-root}}
hdfs://{{host of 3.0/3.0.1 namenode}}/user/as_user/analytic-root
```

3. 在 Ambari 控制台中，停止 Analytic Server 服务。
4. 确保 Analytic Metastore 服务处于运行状态。
5. 从旧安装中收集配置设置。

- a. 将新安装版本中的 configcollector.zip 归档复制到旧安装版本中的 {AS\_ROOT}\tools。
  - b. 解压缩 configcollector.zip 的副本。这将在旧安装版本中创建新的 configcollector 子目录。
  - c. 通过执行 {AS\_ROOT}\tools\configcollector 中的 **configcollector** 脚本，在旧安装版本中运行配置收集器工具。将生成的压缩 (ZIP) 文件复制到托管新安装的服务器。
6. 执行 **migrationtool** 脚本并将配置收集器所创建的压缩文件的路径作为参数进行传递。示例如下所示。  

```
migrationtool.sh /opt/ibm/spss/analyticserver/3.0/ASConfiguration_2.1.0.0.xxx.zip
```
  7. 清除 Zookeeper 状态。在 Zookeeper bin 目录（例如，Hortonworks 上的 /usr/hdp/current/zookeeper-client 或 BigInsights 上的 /usr/iop/current/zookeeper-server）中，运行以下命令。  

```
./zkCli.sh rmr /AnalyticServer
```
  8. 在 Ambari 控制台中，启动 Analytic Server 服务。

注：如果已将 R 配置为与现有 Analytic Server 安装配合使用，那么您将需要遵循相应步骤，将其配置为用于新的 Analytic Server 安装。

---

## 卸载

**要点：**在安装 Essentials for R 时，首先必须运行 remove\_R.sh 脚本。在 Analytic Server 之前未能卸载 Essentials for R 将导致以后无法卸载 Essentials for R。在卸载 Analytic Server 时，将除去 remove\_R.sh 脚本。有关卸载 Essentials for R 的信息，请参阅『卸载 Essentials for R』。

1. 在 Analytic Metastore 主机上，使用以下参数运行 {AS\_ROOT}/bin 目录中的 remove\_as.sh 脚本。

- u** 必需。 Ambari 服务器管理员的用户标识。
- p** 必需。 Ambari 服务器管理员的密码。
- h** 必需。 Ambari 服务器主机名。
- x** 必需。 Ambari 服务器端口。
- l** 可选。 启用安全方式。

示例如下所示：

```
remove_as.sh -u admin -p admin -h one.cluster -x 8081
```

从具有 Ambari 主机 one.cluster 的集群中除去 Analytic Server。

```
remove_as.sh -u admin -p admin -h one.cluster -x 8081 -l
```

以安全方式从具有 Ambari 主机 one.cluster 的集群中除去 Analytic Server。

注：此操作除去 HDFS 上的 Analytic Server 文件夹。

注：该操作不会移除与 Analytic Server 关联的任何 DB2 模式。请参阅 DB2 文档以了解有关手动移除模式的信息

## 卸载 Essentials for R

1. 在 Essentials for R 主机上，在 {AS\_ROOT}/bin 目录中运行具有以下参数的 remove\_r.sh 脚本。

- u** 必需。 Ambari 服务器管理员的用户标识。
- p** 必需。 Ambari 服务器管理员的密码。
- h** 必需。 Ambari 服务器主机名。

**x** 必需。 Ambari 服务器端口。

**l** 可选。 启用安全方式。

示例如下所示：

```
remove_r.sh -u admin -p admin -h one.cluster -x 8081
```

从包含 Ambari 主机 one.cluster 的集群中移除 Essentials for R。

```
remove_r.sh -u admin -p admin -h one.cluster -x 8081 -l
```

从包含 Ambari 主机 one.cluster 的集群中以安全方式移除 Essentials for R。

2. 从 Ambari 服务器服务目录中移除 R 服务目录。 例如，在 BigInsights 4.2 中，ESSENTIALR 目录位于 /var/lib/ambari-server/resources/stacks/BigInsights/4.2/services 中。
3. 在 Ambari 控制台中，验证 Essentials for R 服务是否不再存在。

---

## 第 3 章 Cloudera 安装和配置

---

### Cloudera 概述

Cloudera 是一个开放式源代码 Apache Hadoop 分发版。包含 Apache Hadoop (CDH) 的 Cloudera 分发版针对使用该技术的企业级部署。

Analytic Server 可以在 CDH 平台上运行。CDH 包含 Hadoop 的主要核心元素以及其他面向企业的组件，这些元素提供对大数据集（主要是 MapReduce 和 HDFS）的可靠可伸缩的分布式数据处理，而这些组件提供安全、高可用性以及与硬件和其他软件的集成。

---

### 特定于 Cloudera 的先决条件

除了一般的先决条件，还需查看以下信息。

**服务** 请确保将以下实例安装在每个 Analytic Server 主机上。

- HDFS: Gateway、DataNode 或 NameNode
- Hive: Gateway、Hive Metastore Server 或 HiveServer2
- Yarn: Gateway、ResourceManager 或 NodeManager

以下实例仅在使用其功能时是必需的。

- Accumulo: Gateway
- HBase: Gateway、Master 或 RegionServer

#### 元数据存储库

如果计划将 MySQL 用作 Analytic Server 的元数据存储库，请根据『为 Analytic Server 配置 MySQL』的指示信息进行操作。

### 为 Analytic Server 配置 MySQL

在 Cloudera Manager 中配置 IBM SPSS Analytic Server 要求安装和配置 MySQL 服务器数据库。

1. 从存储 MySQL 数据库的节点上的命令窗口运行以下命令：

```
yum install mysql-server
```

注：对于 SuSE Linux 使用 `zypper install mysql`。

2. 在每个 Cloudera 集群节点上的命令窗口运行以下命令：

```
yum install mysql-connector-java
```

注：对于 SuSE Linux，使用 `sudo zypper install mysql-connector-java`。

3. 决定并记录 Analytic Server 在访问 MySQL 数据库时使用的 Analytic Server 数据库名称、数据库用户名和数据库密码。
4. 根据第 24 页的『在 Cloudera 上安装』中的指示信息安装 Analytic Server。
5. 将 `/opt/cloudera/parcels/AnalyticServer/bin/add_mysql_user.sh` 脚本从 Cloudera 管理的其中一个服务器复制到安装了 MySQL 数据库的节点。使用适合您特定配置的参数运行该脚本。例如：

```
./add_mysql_user.sh -u <database_user_name> -p <database_password> -d  
<database_name>
```

注释：当数据库以安全方式（即设置了 root 用户密码）运行时需要 `a -r <dbRootPassword>` 参数。

当数据库使用非 root 用户名以安全方式运行时需要 `-r <dbUserPassword>` 和 `-t <dbUserName>` 参数。

6. 打开 Cloudera Manager，并导航到 Analytic Server 服务的“配置”选项卡。
  - a. 在 **Analytic Server Metastore 驱动程序类 (jndi.aedb.driver)** 属性中，选择 `com.mysql.jdbc.Driver`
  - b. 对于先前在指定 Analytic Server 配置条目的窗格上记录的 Analytic Server 数据库名称、数据库用户名和数据库密码，必须指定匹配的值。 **Analytic Server Metastore 存储库 URL (jndi.aedb.url)**、**Analytic Server Metastore 用户名 (jndi.aedb.username)** 和 **Analytic Server Metastore 密码 (jndi.aedb.password)** 属性必须更新为与提供给 `add_mysql_user.sh` 命令的值匹配

---

## 在 Cloudera 上安装

以下步骤说明了在 Cloudera Manager 中手动安装 IBM SPSS Analytic Server 的过程。

### 联机安装

1. 在 Cloudera Manager 主集群节点上，下载并运行 Cloudera 自抽取 \*.bin 安装程序。接受许可协议并保持缺省 CSD 安装目录不变，以按照安装提示进行操作。

注：如果从缺省位置变更了 CSD 目录，必须指定其他 CSD 目录。

2. 安装完成后重新启动 Cloudera Manager。
3. 在 Cloudera Manager 界面上（例如，使用缺省登录凭证 `admin/admin` 登录的 `http://${CM_HOST}:7180/cm/login`），刷新**远程包存储库 URL**，并验证这些 URL 是否正确。例如：

```
http://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/cloudera/parcels/latest/  
or  
http://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.0.1.0/cloudera/
```

注：可以更新包更新频率和**远程包存储库 URL** 以满足您的特定需要。

4. 在 Cloudera Manager 刷新包文件（可以通过单击**检查新包**来手动刷新包文件）之后，您会看到 **AnalyticServer** 包状态设置为**远程可用**。
5. 选择**下载 > 分发 > 激活**。 **AnalyticServer** 包状态更新为**已分发和已激活**。
6. 为 Analytic Server 配置 MySQL。
7. 在 Cloudera Manager 中，将 Analytic Server 添加为服务，并决定要放置 Analytic Server 的位置。您需要在“添加服务向导”中提供以下信息：
  - Analytic Server metastore 用户名
  - Analytic Server metastore 密码

“添加服务向导”显示服务创建过程的每个阶段中的总体进度，并在服务成功安装和配置到集群上时提供最终确认消息。

注：在成功安装 Analytic Server 之后，请勿单击 Cloudera Manager 中 Analytic Server 服务页面的“操作”列表上的**创建 Analytic Server Metastore**。创建 metastore 以覆盖现有元数据存储库。

### 脱机安装

脱机安装步骤与联机步骤相同，唯一的差别是您必须手动下载适用于您的特定操作系统的包文件和元数据。

RedHat Linux 需要以下文件：

- AnalyticServer-3.0.1.0-el6.parcel

- AnalyticServer-3.0.1.0-el6.parcel.sha
- manifest.json

or

- AnalyticServer-3.0.1.0-el7.parcel
- AnalyticServer-3.0.1.0-el7.parcel.sha

SuSE Linux 需要以下文件：

- AnalyticServer-3.0.1.0-sles11.parcel
- AnalyticServer-3.0.1.0-sles11.parcel.sha
- manifest.json

1. 在 Cloudera Manager 主集群节点上，下载并运行 Cloudera 自抽取 \*.bin 安装程序。接受许可协议并保持缺省 CSD 安装目录不变，以按照安装提示进行操作。

注：如果与缺省位置不同，必须指定其他 CSD 目录。

2. 将所需的包和元数据文件复制到 Cloudera Manager 主集群节点上的本地 Cloudera repo 路径中。缺省路径是 /opt/cloudera/parcel-repo（该路径可以在 Cloudera Manager 用户界面中配置）。

在 Cloudera Manager 刷新包之后，**AnalyticServer** 包显示为已下载。您可以单击检查新包以强制刷新。

3. 单击分发 > 激活。

此时 **AnalyticServer** 包会显示为已分发和已激活。

## 在 Cloudera 上升级 Analytic Server

如果您已安装 Analytic Server 3.0，那么可以在此基础上将安装升级到 V3.0.1。

1. 在 Cloudera Manager 中，停止然后删除 Analytic Server 服务。
2. 在 Cloudera Manager 中，取消激活 Analytic Server 的先前版本。
3. 有关安装 Analytic Server 3.0.1 的指示信息，请参阅第 18 页的『升级与迁移』中的"联机"和"脱机"部分。
4. 在 Cloudera Manager 中安装并添加 Analytic Server 服务之后，运行刷新 **Analytic Server** 二进制文件。Analytic Server 3.0.1 现在可供使用。

---

## 配置 Cloudera

安装后，您可以通过 Cloudera Manager 有选择地配置和管理 Analytic Server。

注：以下约定用于 Analytic Server 文件路径。

- {AS\_ROOT} 指的是部署 Analytic Server 的位置；例如 /opt/cloudera/parcels/AnalyticServer。
- {AS\_SERVER\_ROOT} 指的是配置文件、日志文件和服务器文件的位置；例如，/opt/cloudera/parcels/AnalyticServer/ae\_wlpserver/usr/servers/aeserver。
- {AS\_HOME} 指的是 HDFS 上 Analytic Server 用作根文件夹的位置；例如，/user/as\_user/analytic-root。

## 安全性

**security\_cfg** 参数用于定义可以作为主体添加至 Analytic Server 系统的用户和组的注册表。

缺省情况下，基本注册表是使用单个用户 `admin` 及密码 `admin` 定义的。您可以通过编辑 `security_cfg` 或将 Kerberos 配置为安全提供程序来更改注册表。`security_cfg` 参数可以在 Analytic Server 服务“配置”选项卡的 **Analytic Server** 高级配置片段部分中找到。

**注：**如果您编辑 `security_cfg` 参数以变更注册表，那么需要将任何新用户作为主体添加至 Analytic Server 系统。请参阅《IBM SPSS Analytic Server 管理员指南》，以获取有关租户管理的详细信息。

## 更改基本注册表

通过基本注册表，您可以在 `security_cfg` 参数中定义用户和组的数据库。

缺省基本注册表类似于以下内容。

```
<basicRegistry id="basic" realm="ibm">
  <user name="admin" password="admin"/>
</basicRegistry>
```

变更后的基本注册表的示例如下所示。

```
<basicRegistry id="basic" realm="ibm">
  <user name="user1" password="{xor}Dz4sLG5tbGs="/>
  <user name="user2" password="Pass"/>
  <user name="user3" password="Pass"/>
  <user name="user4" password="Pass"/>
  <user name="admin" password="{xor}KzosKw="/>
  <group name="Development">
    <member name="user1"/>
    <member name="user2"/>
  </group>
  <group name="QA">
    <member name="user3"/>
    <member name="user4"/>
  </group>
  <group name="ADMIN">
    <member name="user1"/>
    <member name="admin"/>
  </group>
</basicRegistry>
```

可以使用 `{AS_ROOT}/ae_wlpserver/bin` 的 `securityUtility` 工具对密码进行编码，从而对它们的值进行加密。

```
securityUtility encode changeit
  {xor}PDC+MTg6Nis=
```

**注：**请参阅 [http://www-01.ibm.com/support/knowledgecenter/SSD28V\\_8.5.5/com.ibm.websphere.wlp.core.doc/ae/rwlp\\_command\\_securityutil.html](http://www-01.ibm.com/support/knowledgecenter/SSD28V_8.5.5/com.ibm.websphere.wlp.core.doc/ae/rwlp_command_securityutil.html) 以获取 `securityUtility` 工具的详细信息。

**注：**基本注册表用在沙箱环境中，但建议不要将其用于生产环境。

## 配置 LDAP 注册表

LDAP 注册表允许您使用外部 LDAP 服务器（例如，Active Directory 或 OpenLDAP）对用户进行认证。

以下是针对 OpenLDAP 的 `ldapRegistry` 示例。

```
<ldapRegistry
  baseDN="ou=people,dc=aeldap,dc=org"
  ldapType="Custom"
  port="389"
  host="server"
  id="OpenLDAP"
  bindDN="cn=admin,dc=aeldap,dc=org"
```

```

bindPassword="{xor}Dz4sLG5tbGs="
searchTimeout="300000m"
recursiveSearch="true">
<customFilters
  id="customFilters"
  userFilter="(&(uid=%v)(objectClass=inetOrgPerson))"
  groupFilter="(&(cn=%v)(|(objectclass=organizationalUnit)))"
  groupMemberIdMap="posixGroup:memberUid"/>
</ldapRegistry>

```

有关配置的更多示例，请参阅模板文件夹 {AS\_ROOT}/ae\_wlpserver/templates/config。

**注：**Analytic Server 对 LDAP 的支持由 WebSphere Liberty 进行控制。有关更多信息，请参阅在 Liberty 中配置 LDAP 用户注册表。

## 配置从 Analytic Server 到 LDAP 的安全套接字层 (SSL) 连接

1. 以 Analytic Server 用户身份登录每台 Analytic Server 机器，然后为 SSL 证书创建公共目录。

**注：**在 Cloudera 上，Analytic Server 用户始终是 as\_user，这一点无法更改。

2. 将密钥库和信任库文件复制到所有 Analytic Server 机器上的某一公共目录。另外，将 LDAP 客户机 CA 证书添加至信任库。以下是一些指示信息示例。

```

mkdir /home/as_user/security
cd /home/as_user/security
openssl s_client -connect <ldap-hostname>:636 -showcerts > client.cert
$JAVA_HOME/bin/keytool -import -file ./client.cert -alias ldapCA -keystore mytrust.jks
password : changeit

```

**注：**JAVA\_HOME 是用于 Analytic Server 启动的同一 JRE。

3. 可以使用 {AS\_ROOT}/ae\_wlpserver/bin 的 securityUtility 工具对密码进行编码，从而对它们的值进行加密。示例如下所示。

```

securityUtility encode changeit
{xor}Pdc+MTg6Nis=

```

4. 登录 Cloudera Manager 并使用正确的 SSL 配置设置更新 Analytic Server 配置设置 **ssl\_cfg**。示例如下所示。

```

<ssl id="defaultSSLConfig" keyStoreRef="defaultKeyStore" trustStoreRef="defaultTrustStore"
  clientAuthenticationSupported="true"/>
  <keyStore id="defaultKeyStore" location="/home/as_user/security/mykey.jks" type="JKS"
    password="{xor}0zo5PiozKxYdEgwPDaWeDG1uDz4sLCg7"/>
  <keyStore id="defaultTrustStore" location="/home/as_user/security/mytrust.jks" type="JKS"
    password="{xor}Pdc+MTg6Nis="/>

```

**注：**对密钥库文件和信任库文件使用绝对路径。

5. 使用正确的 LDAP 配置设置更新 Analytic Server 配置设置 **security\_cfg**。例如，在 **ldapRegistry** 元素中，将 **sslEnabled** 属性设置为 true 并将 **sslRef** 属性设置为 defaultSSLConfig。

## 配置 Kerberos

Analytic Server 支持在 Cloudera 中使用 Kerberos。

1. 在 Kerberos 用户存储库中，为您计划向其提供对 Analytic Server 的访问权的所有用户创建帐户。

**注：**如果 Analytic Server 安装使用基本注册表，那么它必须包含 Kerberos 用户帐户并使用 "-" 作为密码。示例如下所示。

```

<basicRegistry id="basic" realm="ibm">
  <user name="admin" password="-"/>
  <user name="user1" password="-"/>

```

```

<user name="user2" password="-"/>
<group name="group1">
  <member name="admin"/>
  <member name="user1"/>
  <member name="user2"/>
</group>
<group name="group2">
  <member name="admin"/>
  <member name="user1"/>
</group>
</basicRegistry>

```

2. 在每个 Analytic Server 节点和 Hadoop 节点上，为前一步骤中创建的每个用户创建一个操作系统用户帐户。
  - 确保这些用户的 UID 在所有机器上都匹配。您可以使用 `kinit` 命令登录每个帐户来测试此内容。
  - 确保 UID 符合"用于提交作业的最小用户标识"Yarn 设置。这是 `container-executor.cfg` 中的 `min.user.id` 参数。例如，如果 `min.user.id` 为 1000，那么创建的每个用户帐户都必须具有大于或等于 1000 的 UID。
3. 在 Analytic Server 中，为所有主体在 HDFS 上创建用户主文件夹。例如，如果您将 `testuser1` 添加至 Analytic Server 系统，请在 HDFS 上创建类似于 `/user/testuser1` 的主文件夹，并确保 `testuser1` 具有对此文件夹的读写许可权。
4. [可选] 如果您计划使用 HCatalog 数据源并且 Analytic Server 与 Hive metastore 安装在不同的机器上，那么必须在 HDFS 上模拟 Hive 客户机。
  - a. 在 Cloudera Manager 中，浏览至 HDFS 服务的"配置"选项卡。
 

注：如果尚未设置以下参数，那么可能不会显示在"配置"选项卡上。在此情况下，请运行搜索来查找这些参数。
  - b. 编辑 `hadoop.proxyuser.hive.groups` 参数以使其具有值 `*`，或者编辑包含所有能够登录 Analytic Server 的用户的组。
  - c. 编辑 `hadoop.proxyuser.hive.hosts` 参数以使其具有值 `*`，或者编辑 Hive metastore 和每个 Analytic Server 实例都作为服务安装在其上主机的列表。
  - d. 重新启动 HDFS 服务。

执行这些步骤并且安装了 Analytic Server 步骤后，Analytic Server 将以静默方式自动配置 Kerberos。

## 使用 Kerberos 针对单点登录 (SSO) 配置 HAProxy

1. 按照 HAProxy 文档指南配置并启动 HAProxy：<http://www.haproxy.org/#docs>
2. 针对 HAProxy 主机创建 Kerberos 原则 (`HTTP/<proxyHostname>@<realm>`) 和密钥表文件，其中，`<proxyHostname>` 是 HAProxy 主机的完整名称，而 `<realm>` 是 Kerberos 域。
3. 将密钥表文件作为 `/etc/security/keytabs/spnego_proxy.service.keytab` 复制到每个 Analytic Server 主机
4. 在每个 Analytic Server 主机上更新此文件的许可权。示例如下所示。
 

```

chown root:hadoop /etc/security/keytabs/spnego_proxy.service.keytab
chmod 440 /etc/security/keytabs/spnego_proxy.service.keytab

```
5. 打开 Cloudera Manager，并在 Analytic Server 的 `analyticserver-conf/config.properties` 的 **Analytic Server 高级配置片段（安全值）** 区域中添加或更新以下属性。
 

```

web.authentication.kerberos.keytab=/etc/security/keytabs/spnego_proxy.service.keytab
web.authentication.kerberos.principal=HTTP/<proxy machine full name>@<realm>

```
6. 从 Cloudera Manager 保存配置并重新启动所有 Analytic Server 服务。
7. 指示用户将其浏览器配置为使用 Kerberos。

现在，用户能够使用 Kerberos SSO 登录到 Analytic Server。

## 禁用 Kerberos

1. 在 Ambari 控制台中禁用 Kerberos。
2. 停止 Analytic Server 服务。
3. 从 **analyticserver-conf/config.properties** 的 **Analytic Server** 高级配置片段（安全值）区域中移除以下参数。

```
default.security.provider
hdfs.keytab
hdfs.user
java.security.krb5.conf
jdbc.db.connect.method.kerberos
web.authentication.kerberos.keytab
web.authentication.kerberos.principal
```

4. 单击**保存更改**，然后重新启动 Analytic Server 服务。

## 启用与 Analytic Server 控制台的安全套接字层 (SSL) 连接

缺省情况下，Analytic Server 会生成自签名证书以启用安全套接字层 (SSL)，因此您可以接受自签名证书，通过安全端口来访问 Analytic Server 控制台。要使 HTTPS 访问更安全，您需要安装第三方供应商证书。

要安装第三方供应商证书，请执行下列步骤。

1. 将第三方供应商密钥库和信任库复制到所有 Analytic Server 节点中的同一目录；例如，`/home/as_user/security`。

注：Analytic Server 用户必须具有对此目录的读访问权。

2. 在 Cloudera Manager 中，导航到 Analytic Server 服务的“配置”选项卡。
3. 编辑 **ssl\_cfg** 参数。

```
<ssl id="defaultSSLConfig"
  keyStoreRef="defaultKeyStore"
  trustStoreRef="defaultTrustStore"
  clientAuthenticationSupported="true"/>
<keyStore id="defaultKeyStore"
  location="<KEYSTORE-LOCATION>"
  type="<TYPE>"
  password="<PASSWORD>"/>
<keyStore id="defaultTrustStore"
  location="<TRUSTSTORE-LOCATION>"
  type="<TYPE>"
  password="<PASSWORD>"/>
```

将

- `<KEYSTORE-LOCATION>` 替换为密钥库的绝对位置；例如，`/home/as_user/security/mykey.jks`
- `<TRUSTSTORE-LOCATION>` 替换信任库的绝对位置；例如：`/home/as_user/security/mytrust.jks`
- `<TYPE>` 替换为证书的类型；例如：JKS、PKCS12 等。
- `<PASSWORD>` 替换为使用 Base64 加密格式的加密密码。对于编码，您可以使用 `securityUtility`；例如：`{AS_ROOT}/ae_wlpserver/bin/securityUtility encode <password>`

如果要生成自签名证书，您可以使用 `securityUtility`；例如：`{AS_ROOT}/ae_wlpserver/bin/securityUtility createSSLCertificate --server=myserver --password=mypassword --validity=365 --subject=CN=mycompany,O=myOrg,C=myCountry`。有关 `securityUtility` 和其他 SSL 设置的更多信息，请参阅 `WebSphere Liberty` 概要文件文档。

4. 单击**保存更改**，然后重新启动 Analytic Server 服务。

## 启用对 **Essentials for R** 的支持

Analytic Server 支持对 R 模型进行评分，并支持运行 R 脚本。

要在 Cloudera Manager 中成功安装 Analytic Server 之后安装 Essentials for R：

1. 下载用于 IBM SPSS Modeler Essentials for R RPM 的自解压缩归档 (BIN)。Essentials for R 可从以下地址下载：<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>。选择特定于堆栈、堆栈版本及硬件体系结构的文件。
2. 在 Cloudera Manager 服务器主机上以 root 或 sudo 用户身份运行自抽取归档。以下包必须安装到已配置的存储库或可从已配置的存储库中使用：
  - Red Hat Linux: gcc-gfortran、zip 和 gcc-c++
  - SUSE Linux: gcc-fortran、zip 和 gcc-c++
3. 自抽取安装程序执行以下任务：
  - a. 显示必需许可证并提示安装程序接受这些许可证。
  - b. 提示安装程序输入 R 源位置，或者使用缺省位置继续操作。安装的缺省 R 版本是 3.1.0。要安装其他版本：
    - 联机安装：提供指向必需 R 版本归档的 URL。例如，<https://cran.r-project.org/src/base/R-2/R-2.15.3.tar.gz> 表示 R 2.15.3。
    - 脱机安装：下载必需 R 版本归档，然后将该归档复制到 Cloudera Manager 服务器主机。请勿重命名该归档（缺省情况下，其名称为 R-x.x.x.tar.gz）。提供指向所复制的 R 归档的 URL，如下所示：`file://<R_archive_directory>/R-x.x.x.tar.gz`。如果下载了 R-2.15.3.tar.gz 归档并将其复制到 /root，URL 为 `file:///root/R-2.15.3.tar.gz`。

注：其他 R 版本可以在 <https://cran.r-project.org/src/base/> 中找到。
  - c. 安装 R 所需的软件包。
  - d. 下载并安装 R 以及 Essentials for R 插件。
  - e. 创建包和 parcel.sha 文件，并将它们复制到 /opt/cloudera/parcel-repo。如果位置已发生更改，请输入正确的位置。
4. 安装完成后，在 Cloudera Manager 中分发并激活 **Essentials for R** 包（单击**检查新包**以刷新包列表）。
5. 如果 Analytic Server 服务已安装：
  - a. 停止服务。
  - b. 刷新 Analytic Server 二进制文件。
  - c. 启动该服务以完成 Essentials for R 安装。
6. 如果未安装 Analytic Server 服务，那么继续其安装。

注：所有 Analytic Server 主机都必须安装相应的归档（zip 和 unzip）软件包。

## 启用关系数据库源

如果您在每个 Analytic Server 主机的共享目录中提供 JDBC 驱动程序，那么 Analytic Server 可以使用关系数据库源。缺省情况下，此目录为 /usr/share/jdbc。

要更改共享目录，请执行下列步骤。

1. 在 Cloudera Manager 中，导航到 Analytic Server 服务的"配置"选项卡。

2. 在 `jdbc.drivers.location` 中, 指定 JDBC 驱动程序的共享目录的路径。
3. 单击保存更改。
4. 从操作下拉菜单中选择停止以停止 Analytic Server 服务。
5. 从操作下拉菜单中选择刷新 **Analytic Server** 二进制文件。
6. 从操作下拉菜单中选择启动以启动 Analytic Server 服务。

表 3. 受支持的数据库

数据库	受支持的版本	JDBC 驱动程序 JAR	供应商
Amazon Redshift	8.0.2 或更高版本	RedshiftJDBC41-1.1.6.1006.jar 或更高版本	Amazon
DashDB	Bluemix 服务	db2jcc.jar	IBM
DB2 for Linux、UNIX 和 Windows	10.5、10.1 和 9.7	db2jcc.jar	IBM
DB2 z/OS	11 和 10	db2jcc.jar 和 db2_license_cisuz.jar	IBM
Greenplum	5 和 4.2.x	postgresql.jar	Greenplum
Hive	1.1、1.2	hive-jdbc-*.jar	Apache
Netezza	7 和 6.x	nzjdbc.jar	IBM
Oracle	12c 和 11g R2 (11.2)	ojdbc6.jar 和 orai18n.jar	Oracle
SQL Server	2014、2012 和 2008 R2	sqljdbc4.jar	Microsoft
Sybase IQ	16.x、15.4 和 15.2	jconnect70.jar	Sybase
Teradata	14、14.1 和 15	tdgssconfig.jar 和 terajdbc4.jar	Teradata

## 启用 HCatalog 数据源

Analytic Server 通过 Hive/HCatalog 提供对许多数据源的支持。一些源需要手动配置步骤。

1. 收集必需的 JAR 文件以启用数据源。请参阅下列各节以获取详细信息。
2. 将这些 JAR 文件添加至 `{HIVE_HOME}/auxlib` 目录及每个 Analytic Server 节点的 `/usr/share/hive` 目录。
3. 重新启动 Hive Metastore 服务。
4. 重新启动 Analytic Server 服务的每个实例。

## NoSQL 数据库

Analytic Server 支持任何 NoSQL 数据库, 可以从供应商处获取用于此数据库的 Hive 存储处理程序。

不需要任何额外步骤即可启用对 Apache HBase 及 Apache Accumulo 的支持。

对于其他 NoSQL 数据库, 请联系联系人供应商并获取存储处理程序及相关 JAR。

## 基于文件的 Hive 表

Analytic Server 支持任何基于文件的 Hive 表, 有一个内置或定制 Hive SerDe (序列化器-反序列化器) 可用于此表。

用于处理 XML 文件的 Hive XML SerDe 位于 Maven Central Repository 中，网址为 <http://search.maven.org/#search%7Cga%7C1%7Chivexmlserde>。

## Apache Spark

如果要使用 Spark (V1.5 或更高版本) 和 HCatalog 输入数据源，必须手动添加 `spark.version=1.5.0` 属性。

1. 打开 Cloudera Manager，并在 Analytic Server 的 **analyticserver-conf/config.properties** 的高级配置片段（安全值）区域中添加或更新以下属性。

```
spark.version=1.5.0
```

2. 从 Cloudera Manager 保存配置并重新启动所有 Analytic Server 服务。

## 更改 Analytic Server 所使用的端口

缺省情况下，Analytic Server 对 HTTP 和 HTTPS 分别使用 9080 和 9443 端口。要更改端口设置，请执行下列步骤。

1. 在 Cloudera Manager 中，导航到 Analytic Server 服务的“配置”选项卡。
2. 在 **http.port** 和 **https.port** 中分别指定所需的 HTTP 和 HTTPS 端口。

注：您可能需要选择“过滤”部分中的端口和地址类别才能看到这两个参数。

3. 单击保存更改。
4. 重新启动 Analytic Server 服务。

## 高可用性 Analytic Server

通过将 Analytic Server 作为服务添加至集群中的多个节点，您可以实现它的高可用性。

1. 在 Cloudera Manager 中，导航到 Analytic Server 服务的“实例”选项卡。
2. 单击添加角色实例并选择要将 Analytic Server 添加为服务的主机。

## 针对小型数据优化 JVM 选项

您可以编辑 JVM 属性，以便在运行小型 (M3R) 作业时优化系统。

在 Cloudera Manager 中，请查看 Analytic Server 服务中“配置”选项卡上的 **JVM 选项 (jvm.options)** 控件。修改下列参数会设置托管 Analytic Server 的服务器上运行的作业的堆大小；即，并非 Hadoop。运行较小的 (M3R) 作业时，这一点很重要，并且您可能需要对这些值进行试验以优化系统。

```
-Xms512M  
-Xmx2048M
```

---

## 迁移

通过 Analytic Server，您可以将数据和配置设置从现有 Analytic Server 迁移到新安装。

### 升级到新版本的 Analytic Server

如果您已安装 Analytic Server 2.0/2.1 并已购买更新的版本，那么可以将 2.0/2.1 配置设置迁移到新安装。

**限制：**如果您已安装低于 2.0 的版本，那么必须先从较低版本迁移到 V2.0/V2.1，然后再从 V2.0/V2.1 迁移到更新的版本。

**限制：**2.0/2.1 和新安装无法在同一 Hadoop 集群中共存。如果您将新安装配置为使用 2.0/2.1 安装所用的 Hadoop 集群，那么 2.0/2.1 安装将不再起作用。

## 2.1 到更新版本的迁移步骤

1. 根据第 24 页的『在 Cloudera 上安装』中的指示信息，安装新的 Analytic Server。
2. 将分析根目录从旧安装复制到新安装。
  - a. 如果您不确定分析根目录的位置，请运行 `hadoop -fs ls`。分析根目录的路径的格式为 `/user/aeuser/analytic-root`，其中 `aeuser` 是拥有分析根目录的用户标识。
  - b. 将分析根目录的所有权从 `aeuser` 更改为 `as_user`

```
hadoop dfs -chown -R {as_user:{group}} {path to 2.1 analytic-root}
```

注：如果您打算在迁移后使用现有的 Analytic Server 安装，请在 HDFS 生成 `analytic-root` 目录的副本，然后更改该目录副本的所有权。
  - c. 以 `as_user` 身份登录新 Analytic Server 安装版本的主机。如果 `/user/as_user/analytic-root` 目录存在，请将其删除。
  - d. 运行下列复制脚本。

```
hadoop distcp hftp://{host of 2.1 namenode}:50070/{path to 2.1 analytic-root}
hdfs://{host of 3.0 namenode}/user/as_user/analytic-root
```
3. 在 Cloudera Manager 中，停止 Analytic Server 服务。
4. 从旧安装中收集配置设置。
  - a. 将新安装版本中的 `configcollector.zip` 归档复制到旧安装版本中的 `{AS_ROOT}\tools`。
  - b. 解压缩 `configcollector.zip` 的副本。这将在旧安装版本中创建新的 `configcollector` 子目录。
  - c. 通过执行 `{AS_ROOT}\tools\configcollector` 中的 **configcollector** 脚本，在旧安装版本中运行配置收集器工具。将生成的压缩 (ZIP) 文件复制到托管新安装的服务器。
5. 执行 **migrationtool** 脚本并将配置收集器所创建的压缩文件的路径作为参数进行传递。示例如下所示。

```
migrationtool.sh /opt/ibm/spss/analyticserver/3.0/ASConfiguration_2.1.0.0.xxx.zip
```
6. 清除 Zookeeper 状态。在 Zookeeper bin 目录中（例如 Cloudera 上的 `/opt/cloudera/parcels/CDH-5.4...../lib/zookeeper/bin`），运行以下命令。

```
./zkCli.sh rmr /AnalyticServer
```
7. 在 Cloudera Manager 中，启动 Analytic Server 服务。

注：如果已将 R 配置为与现有 Analytic Server 安装配合使用，那么您将需要遵循相应步骤，将其配置为用于新的 Analytic Server 安装。

---

## 在 Cloudera 上卸载 Analytic Server

Cloudera 自动处理卸载 Analytic Server 服务和包所需的大部分步骤。

以下是从 Cloudera 环境中清除 Analytic Server 所必需的步骤：

1. 停止并删除 Analytic Server 服务。
2. 取消激活并从主机移除 Analytic Server 包。
3. 删除 HDFS 中的 Analytic Server 用户目录。缺省位置是 `/user/as_user/analytic-root`。
4. 删除 Analytic Server 使用的数据库或模式。



---

## 第 4 章 MapR 安装和配置

---

### MapR 概述

MapR 是针对 Apache Hadoop 的完整分发版，其中打包了来自 Hadoop 生态系统的数十个项目以提供广泛的大数据功能。

无法在服务器集群外部访问 MapR 文件系统。因此，IBM SPSS Analytic Server 必须部署到 MapR 集群节点中。在该部署方案中，Analytic Server 必须由有权访问 MapR 文件系统以及向 yarn 提交作业以部署到 Analytic Server（以 <as\_user> 身份）的用户运行。

---

### 在 MapR 上安装 Analytic Server

以下步骤详细描述了在 MapR 集群上手动安装 IBM SPSS Analytic Server 的过程。

1. 以 root 或 sudo 用户身份运行 Analytic Server 安装程序 (spss\_as-3.0.1.0-mapr5.1-1x86-64\_en.bin)。根据安装提示进行操作以接受许可证，并选择联机或脱机安装 Analytic Server。
  - a. 如果托管 Analytic Server 的服务器具有与 http://ibm-open-platform.ibm.com 的因特网连接，请选择联机选项。安装程序将自动安装 Analytic Server。
  - b. 如果托管 Analytic Server 的服务器不具有与 http://ibm-open-platform.ibm.com 的因特网连接，请选择脱机选项。在能够访问该 URL 的另一服务器上运行安装程序，并选择脱机安装 Analytic Server。安装程序会自动下载 RPM 软件包。

2. 查找并运行 Analytic Server 的 RPM:

```
rpm -ivh IBM-SPSS-AnalyticServer-3.0.1.0-1.x86_64.rpm
```

对于联机和脱机安装方式，Analytic Server 安装到 /opt/ibm/spss/analyticserver/3.0（作为 <as\_installation\_path>）。

3. 将安装路径中的所有文件更改为运行 Analytic Server 的用户:

```
chown -R <as_user> <as_installation_path>
```

将用户切换为 <as\_user>; 所有后续步骤均使用 <as\_user>。

4. 配置 HTTP 属性。在路径 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver 中创建名为 http\_endpoint.xml 的文件，然后将以下行添加到该文件:

```
<server>
  <httpEndpoint host="*" id="defaultHttpEndpoint" httpPort="<http_port>" httpsPort="<https_port>" onError="FAIL"/>
</server>
```

<http\_port> 和 <https\_port> 是 Analytic Server 通过 HTTP 和 HTTPS 协议使用的端口。将它们替换为任何可用的端口。

5. 添加用户和组。在路径 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver 中创建名为 security\_cfg.xml 的文件，然后将以下行添加到该文件:

```
<server>
  <basicRegistry id="basic" realm="ibm">
    <user name="admin" password="test"/>
  </basicRegistry>
</server>
```

在缺省状态下，XML 文件仅包含管理用户。您必须手动将其他用户和组添加到 <basicRegistry> 设置中，或者将设置更改为 ldapRegistry。

6. 设置元数据数据库。 Analytic Server 支持 DB2 和 MYSQL 数据库。

- a. 配置数据库用户。 当使用 MySQL 数据库时，在 MySQL shell 中运行以下 SQL 脚本：

```
DROP DATABASE IF EXISTS <db_name>;
CREATE DATABASE <db_name> DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_bin;
CREATE USER '<db_username>@%' IDENTIFIED BY '<db_password>';
CREATE USER '<db_username>@localhost' IDENTIFIED BY '<db_password>';
GRANT ALL PRIVILEGES ON *.* TO '<db_username>@%';
GRANT ALL PRIVILEGES ON *.* TO '<db_username>@localhost';
```

- b. 对密码加密。 数据库用户的密码必须先加密，然后才能将其传递给 Analytic Server。 运行以下命令：

```
java -Duser.language=en -cp <as_installation_path>/ae_wlpserver/usr/servers/aeserver/apps/AE_BOOT.war/WEB-INF/lib/*
com.spss.ae.encryption.provider.EncryptKeystorePassword <db_password>
```

注：当直接在 Linux shell 中运行该命令时，字符 \* 可能需要转义为 \\*。

命令输出内容为：The encrypted password is '<encrypted\_db\_password>'。 记录加密的数据库密码。

- c. 删除文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/configuration/config.properties（如果存在），并使用相同名称创建新文件。 在使用 DB2 数据库时，更改以下属性：

```
jndi.aedb=jdbc/aeds
jndi.aedb.url=jdbc:db2://<db_host>:<db_port>/<db_name>:currentSchema=<db_schema_name>;
jndi.aedb.driver=com.ibm.db2.jcc.DB2Driver
jndi.aedb.username=<db_username>
jndi.aedb.password=<encrypted_db_password>
```

如果 <db\_schema\_name> 模式不存在，那么用户 <db\_username> 必须具有创建该模式的隐式许可权。 在使用 MySQL 数据库时，更改以下属性：

```
jndi.aedb=jdbc/aeds
jndi.aedb.url=jdbc:mysql://<db_host>:<db_port>/<db_name>?createDatabaseIfNotExist=true
jndi.aedb.driver=com.mysql.jdbc.Driver
jndi.aedb.username=<db_username>
jndi.aedb.password=<encrypted_db_password>
```

- d. 在使用 MySQL 数据库时必须安装 MySQL JDBC 驱动程序。 运行以下命令：

```
yum install mysql-connector-java
```

- e. 运行以下命令以创建所需表：

```
cd <as_installation_path>/ae_wlpserver/usr/servers/aeserver/sql/<db_type>
java -Xmx128m -Xms128m -cp <as_installation_path>/ae_wlpserver/usr/servers/aeserver/apps/AE_BOOT.war/WEB-INF/lib/*:/usr/share/java/*
com.spss.ae.dbscript.ScriptRunner ../../configuration/config.properties schema.sql true
```

<db\_type> 为 db2 或 mysql，具体取决于所使用的数据库。

注：当使用带有 MYISAM 引擎的 MySQL 时，第二个命令报告以下错误消息，可以安全地忽略此消息：

```
Error executing: set global innodb_large_prefix=ON
java.sql.SQLException: Unknown system variable 'innodb_large_prefix'
Error executing: set global innodb_file_format=BARRACUDA
java.sql.SQLException: Unknown system variable 'innodb_file_format'
Error executing: set global innodb_file_format_max=BARRACUDA
java.sql.SQLException: Unknown system variable 'innodb_file_format_max'
Error executing: set global innodb_file_per_table=TRUE
java.sql.SQLException: Variable 'innodb_file_per_table' is a read only variable
```

7. 运行以下命令以对 cf 库解包。

```
cd <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration
unzip cf.zip
```

8. 通过在路径 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver 中创建名为 private\_library.xml 的文件来配置 JAAS 登录模块类路径，并在该文件中输入以下信息：

```
<server>
<library id="maprLib">
  <fileset dir="{wlp.install.dir}/usr/servers/aeserver/apps/AE_BOOT.war/WEB-INF/lib" includes="*.jar"/>
  <fileset dir="/usr/share/java" includes="*.jar"/>
  <folder dir="/opt/mapr/hadoop/hadoop-2.7.0/etc/hadoop"/>
  <fileset dir="/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/common" includes="*.jar"/>
  <fileset dir="/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/common/lib" includes="*.jar"/>
  <fileset dir="/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/hdfs" includes="*.jar"/>
</library>
</server>
```

```

<fileset dir="/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/hdfs/lib" includes="*.jar"/>
<fileset dir="/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/yarn" includes="*.jar"/>
<fileset dir="/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/yarn/lib" includes="*.jar"/>
<fileset dir="/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/mapreduce" includes="*.jar"/>
<fileset dir="/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/lib" includes="*.jar"/>
</library>
<jaasLoginModule id="maprLoginModule1" className="org.apache.hadoop.security.login.GenericOSLoginModule"
  controlFlag="REQUIRED" libraryRef="maprLib"></jaasLoginModule>
<jaasLoginModule id="maprLoginModule2" className="org.apache.hadoop.security.login.HadoopLoginModule"
  controlFlag="REQUIRED" libraryRef="maprLib"></jaasLoginModule>
<jaasLoginContextEntry id="hadoop_simple" name="hadoop_simple" loginModuleRef="maprLoginModule1,maprLoginModule2" />
<application context-root="/analyticserver" id="AS_BOOT" location="AE_BOOT.war" name="AS_BOOT" type="war">
  <classloader commonLibraryRef="maprLib"></classloader>
</application>
<application id="help" location="help.war" name="help" type="war" context-root="/analyticserver/help"/>
</server>

```

注：上述示例用于配置 `hadoop_simple` login 模块。如果 MapR 使用其他登录模块，必须更改此配置。

9. 验证文件 `ASModules.xml` 是否存在于路径 `<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/` 中。如果该文件不存在，将文件 `ASModules.xml.template`（位于同一路径中）重命名为 `ASModules.xml`
10. 通过在文件 `<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/config.properties` 中添加以下属性来配置集群信息。

```

ae.cluster.zookeeper.connect.string=
ae.cluster.member.name=
ae.cluster.collective.name=mapr_5.1

```

`ae.cluster.zookeeper.connect.string` 属性是以逗号分隔的 zookeeper 节点列表。该属性可以共享 MapR 使用的 zookeeper 集群。`ae.cluster.member.name` 是托管 Analytic Server 的节点的主机名。

11. 打开文件 `<as_installation_path>/ae_wlpserver/usr/servers/aeserver/server.env` 并向该文件添加以下行：

```

JAVA_HOME=<java_home>
PATH=<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/linux/lib_64:<java_home>/jre/lib/amd64:/usr/sbin:/usr/bin:/sbin:/bin
IBM_SPSS_AS_NATIVE_PATH=<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/linux/lib_64
LD_LIBRARY_PATH=<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/linux/lib_64:<java_home>/jre/lib/amd64:/opt/mapr/hadoop/hadoop-2.7.0/lib/native

```

将 `<as_installation_path>` 和 `<java_home>` 替换为实际安装路径和 Java 主路径。

12. 通过打开文件 `<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/config.properties` 并添加以下行来编辑分析根目录：

```
distrib.fs.root=<analytic_root>
```

`<analytic_root>` 是托管必要 Analytic Server 远程文件的 MapR 文件系统中的路径。建议的路径是 `/user/<as_user>/analytic-root`。

13. 通过打开文件 `<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/config.properties` 并添加以下行来设置管理员用户：

```
admin.username=admin
```

该值必须是 Analytic Server 管理员用户名，并且必须是 `security_cfg.xml` 文件中配置的其中一个用户。

14. 通过在文件 `<as_installation_path>/bin/hdfsUpdate.sh` 的第 69 行添加以下行，将 Analytic Server 依赖关系上载到 MapR 文件系统：

```
JAVA_CLASS_PATH=`hadoop classpath`:$JAVA_CLASS_PATH
```

运行以下命令来创建 `<analytic_root>`：

```
cd <as_installation_path>/bin
./hdfsUpdate.sh
```

<as\_user> 必须具有对 <analytic\_root> 父目录的写许可权。

## 15. 启动并停止 Analytic Server。

### a. 运行以下命令以启动 Analytic Server:

```
cd <as_installation_path>/ae_wlpserver/bin
./server start aeserver
```

### b. 运行以下命令以停止 Analytic Server:

```
cd <as_installation_path>/ae_wlpserver/bin
./server stop aeserver
```

---

## 配置 MapR

安装之后，您可以选择配置和管理 Analytic Server MapR 功能部件。

## 启用数据库回送

数据库回送是从数据库读取数据并直接处理数据的做法。

IBM SPSS Analytic Server 针对以下数据库支持回送:

- DashDB
- DB2
- DB2 for Z
- Hive
- MySQL
- Netezza
- Oracle
- PostgreSQL
- Redshift
- SQL Server
- Sybase IQ
- Teradata

使用以下步骤来启用数据库回送。

1. 将相应的 JDBC 驱动程序 JAR 文件复制到 <as\_installation\_path>/jdbc。
2. 打开文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/private\_library.xml，找到标识为 maprLib 的标记库，并在该标记中添加以下行:

```
<fileset dir="<as_installation_path>/jdbc" includes="*.jar"/>
```

3. 运行以下命令:

```
cd <as_installation_path>/jdbc
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
```

4. 重新启动 Analytic Server。

## 启用 Apache Hive

Apache Hive 是在 Hadoop 上构建的数据仓库基础结构，用于提供数据摘要、查询和分析。

注：Hive 必须配置为将 MySQL 用作 metastore。在托管 IBM SPSS Analytic Server 的节点中存在的 hive-site.xml 文件应与运行 Hive metastore 的节点中的文件相同。

要在成功安装 MapR 之后启用 Apache Hive 支持：

1. 运行以下命令以将 Hive 和 hcatalog 依赖关系上载到 MapR 文件系统：

```
cd /opt/mapr/hive/hive-1.2/lib
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
cd /opt/mapr/hive/hive-1.2/hcatalog/share/hcatalog
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
```

<as\_analytic\_root> 是第 35 页的『在 MapR 上安装 Analytic Server』中定义的分析根路径。

2. 打开文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/private\_library.xml，找到标识为 maprLib 的标记库，并在该标记中添加以下行：

```
<fileset dir="/opt/mapr/hive/hive-1.2/lib" includes="*.jar"/>
<fileset dir="/opt/mapr/hive/hive-1.2/hcatalog/share/hcatalog" includes="*.jar"/>
```

3. 运行以下命令并创建 Hive 和 hcatalog 配置文件链接：

```
mkdir <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/hive-conf
ln -s /opt/mapr/hive/hive-1.2/conf/* <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/hive-conf
```

4. 如果 Hive 的 auxlib 中有额外 jar 文件，将以下行添加到 private\_library.xml 文件：

```
<fileset dir="/opt/mapr/hive/hive-1.2/auxlib" includes="*.jar"/>
```

在添加上述行之后运行以下命令：

```
cd /opt/mapr/hive/hive-1.2/auxlib
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
```

5. 重新启动 Analytic Server。

## 以 HTTP 方式运行 Hive

缺省情况下，Hive 以二进制方式（TCP 方式）运行。要以 HTTP 方式运行 Hive，必须更新以下 Hive 配置属性（特别是 hive.server2.transport.mode 属性）。

注：有关每个属性的更多信息，请参阅 <https://cwiki.apache.org/confluence/display/Hive/Configuration+Properties#ConfigurationProperties-hive.server2.transport.modeHive> 配置属性。

表 4. HTTP 方式下的 Hive 属性

属性名称	缺省值	描述
hive.server2.transport.mode	binary	服务器传输方式。值可能是 binary 或 http。设置为 http 以启用 HTTP 传输方式。
hive.server2.thrift.http.port	10001	HTTP 方式下的端口号。
hive.server2.thrift.http.path	cliservice	HTTP 方式下的 URL 端点的路径部分。
hive.server2.thrift.http.min.worker.threads	5	HTTP 方式下服务器池中工作程序线程的最小数目。
hive.server2.thrift.http.max.worker.threads	500	HTTP 方式下服务器池中工作程序线程的最大数目。

注：必须在更新属性后重新启动 Hive。

## 启用 Apache HBase

Apache HBase 是以 Java 编写的开放式源代码、分布式的非关系数据库。它作为 Apache 软件基金会的 Apache Hadoop 项目的一部分开发，在 HDFS（Hadoop 分布式文件系统）上运行。

要在成功安装 MapR 之后启用 Apache HBase 支持：

1. 将 HBase 依赖关系上载到 MapR 文件系统并运行以下命令：

```
cd /opt/mapr/hbase/hbase-0.98.12/lib
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
```

<as\_analytic\_root> 是第 35 页的『在 MapR 上安装 Analytic Server』中定义的分析根路径。

2. 打开文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/private\_library.xml，找到标识为 maprLib 的标记库，并在该标记中添加以下行：

```
<fileset dir="/opt/mapr/hbase/hbase-0.98.12/lib" includes="*.jar"/>
```

3. 运行以下命令以创建 HBase 和 hcatalog 配置文件链接：

```
mkdir <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/hbase-conf  
ln -s /opt/mapr/hbase/hbase-0.98.12/conf/* <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/hbase-conf
```

4. 重新启动 IBM SPSS Analytic Server。

## 启用 Apache Spark

Apache Spark 是针对灵活内存中数据处理以用于批量、实时和高级分析的开放式标准。

要在成功安装 MapR 之后启用 Apache Spark 支持：

1. 将文件 spark-assembly-1.4.1-hadoop2.5.1-mapr-1501.jar 从 /opt/mapr/spark/spark-1.4.1/lib 复制到 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/modules/spark/。
2. 将 Spark 依赖关系上载到 MapR 文件系统并运行以下命令：

```
cd <as_installation_path>/ae_wlpserver/usr/servers/aeserver/modules/spark/  
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
```

<as\_analytic\_root> 是第 35 页的『在 MapR 上安装 Analytic Server』中定义的分析根路径。

3. 打开文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/private\_library.xml，找到标识为 maprLib 的标记库，并在该标记中添加以下行：

```
<fileset dir="/opt/mapr/spark/spark-1.4.1/lib" includes="spark-assembly-*.jar"/>
```

4. 运行以下命令并创建 Spark 配置文件链接：

```
mkdir <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/spark-conf  
ln -s /opt/mapr/spark/spark-1.4.1/conf/* <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/spark-conf
```

5. 将以下行添加到文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/server.env：

```
SPARK_HOME=/opt/mapr/spark/spark-1.4.1
```

6. 将以下行添加到文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/configuration/config.properties 中：

```
spark.executor.extraLibraryPath=/opt/mapr/hadoop/hadoop-2.7.0/lib/native
```

7. 重新启动 IBM SPSS Analytic Server。

8. 要启用 PySpark 功能部件，将以下行添加到 yarn-env.sh 文件，然后重新启动 ResourceManagers 和 NodeManagers：

```
export SPARK_HOME=/opt/mapr/spark/spark-1.4.1
```

## 启用功能部件标志

功能部件标志提供启用和禁用特定应用程序功能部件的功能。

要在成功安装 MapR 之后启用功能部件标志支持：

1. 将以下行添加到文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/configuration/config.properties 中：

```
load.feature.flags.on.msg=true
```

2. 重新启动 IBM SPSS Analytic Server。

## 启用 R

R 是用于统计计算和图形的语言和环境。

要在成功安装 MapR 之后启用 R 支持：

**注：**必须先安装以下软件包，然后才能在托管节点管理器和 IBM SPSS Analytic Server 的所有集群节点上运行安装程序：

```
gcc-gfortran
libgfortran
gcc-c++
```

1. 在托管节点管理器和 Analytic Server 的所有集群节点上运行安装程序 `spss_er-8.3.0.0-mapr5.1-1x86_64_en.bin`。运行该安装程序的用户必须具有对 R 和 Analytic Server 安装路径的写许可权。
2. 通过接受许可协议来根据安装指示信息进行操作，并输入所需的信息。如果 Analytic Server 安装在安装服务器上，那么在提示时选择是并输入 `<as_installation_path>`。如果 Analytic Server 没有安装在安装服务器上，那么在提示时选择否。
3. 如果安装了 Analytic Server，Essentials for R 自动安装在 Analytic Server 安装路径中。
  - 如果未安装 Analytic Server，Essentials for R 将安装在路径 `<installer_path>/IBM_SPSS_ModelerEssentialsR/linux` 中。
  - 如果稍后安装 Analytic Server，使用以下命令将 Essentials for R 复制到安装了 Analytic Server 的 Analytic Server 配置路径。

```
cp -r <installer_path>/IBM_SPSS_ModelerEssentialsR/linux <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration
```

4. 删除路径 `<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration` 中的 `cf.zip` 文件并使用以下命令生成新文件：

```
cd <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration
zip -r cf.zip linux
```

5. 运行以下命令：

```
cd <as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration
hadoop fs -rm <as_analytic_root>/cluster1/configuration/cf.zip
hadoop fs -put cf.zip <as_analytic_root>/cluster1/configuration/
```

6. 重新启动 Analytic Server。

## 启用 LZO

LZO 是无损数据压缩库，优先选择速度而非压缩比率。MapR 必须手动配置为提供 LZO 支持。

以下站点提供 LZO 安装和配置指示信息：<https://github.com/twitter/hadoop-lzo>。

以下步骤详细描述了将 LZO 库导入 MapR 的过程。

1. 将 `hadoop-lzo-<version>.jar` 文件复制到 Hadoop 类路径。建议的路径是 `/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/lib`。
2. 将本机文件 `libgplcompression.so` 和 `liblzo2.so.2` 复制到 `/opt/mapr/hadoop/hadoop-2.7.0/lib/native`，并将以下属性添加到 `core-site.xml` 文件：

```
<property>
  <name>io.compression.codecs</name>
  <value>org.apache.hadoop.io.compress.GzipCodec,org.apache.hadoop.io.compress.DefaultCodec,com.hadoop.compression.lzo.LzoCodec,com.hadoop.compression.lzo.LzopCodec,org.apache.hadoop.io.compress.BZip2Codec</value>
</property>
<property>
  <name>io.compression.codec.lzo.class</name>
  <value>com.hadoop.compression.lzo.LzoCodec</value>
</property>
```

3. 打开文件 `<as_installation_path>/ae_wlpserver/usr/servers/aeserver/server.env` 并将 `<lzo_native_path>` 添加到 `LD_LIBRARY_PATH` 参数。 `<lzo_native_path>` 是包含 Hadoop-LZO 本机库的文件夹。

```
LD_LIBRARY_PATH=<as_installation_path>/ae_wlpserver/usr/servers/aeserver/configuration/linux/lib_64:<java_home>/jre/lib/amd64:/opt/mapr/hadoop/hadoop-2.7.0/lib/native:<lzo_native_path>
```

4. 重新启动 IBM SPSS Analytic Server。

## 为 MapR 设置 IBM SPSS Analytic Server 集群

使用以下步骤设置 IBM SPSS Analytic Server 集群环境以获取 MapR 支持。

1. 将以下行添加到文件 <as\_installation\_path>/ae\_wlpserver/usr/servers/aeserver/configuration/config.properties 中。

```
enable.resume=true
```
2. 将安装路径复制到其他集群节点，并将 config.properties 文件中的 ae.cluster.member.name 属性更改为正确主机名。
3. 启动所有集群节点。

---

## 卸载 MapR

以下步骤说明了卸载 MapR 的过程：

1. 停止 IBM SPSS Analytic Server。
2. 删除元数据数据库。
  - a. 运行以下命令：

```
cd <as_installation_path>/ae_wlpserver/usr/servers/aeserver/sql/<db_type>
java -Xmx128m -Xms128m -cp <as_installation_path>/ae_wlpserver/usr/Servers/aeserver/apps/AE_BOOT.war/WEB-INF/lib/*:/usr/share/java/*
com.spss.ae.dbscript.ScriptRunner ../../configuration/config.properties drop.sql true
```

- b. 运行以下 SQL 语句以删除数据库：

```
drop database <db_name>
```

3. 卸载 RPM 软件包：

```
rpm -e IBM-SPSS-AnalyticServer-3.0.1.0-1.x86_64
```

4. 删除安装路径：

```
rm -r <as_installation_path>
```

5. 删除分析根目录：

```
hadoop fs -rm -r <analytic-root>
```

6. 删除 zookeeper 数据：

```
/opt/mapr/zookeeper/zookeeper-3.4.5/bin/zkCli.sh -server <zookeeper_host>:<zookeeper_port>
rmr /AnalyticServer
```

---

## 在 MapR 上迁移 IBM SPSS Analytic Server

可以在 MapR 上迁移 IBM SPSS Analytic Server。

使用以下步骤在 MapR 上将 IBM SPSS Analytic Server 2.0 或 2.1 迁移到 V3.0.1。

1. 按照第 35 页的『在 MapR 上安装 Analytic Server』中的安装指示信息在 MapR 集群上安装 Analytic Server 3.0.1。
2. 复制分析根目录。

注：如果分析根目录未更改，可以忽略此步骤。

- 如果 Analytic Server V2.0/V2.1 和 V3.0.1 的分析根目录位于同一个 MapR 集群上，那么请在其中一个数据节点上运行以下命令：

```
hadoop fs -cp <old_analytic_root>/analytic-workspace/* <new_analytic_root>/analytic-workspace
```

- 当 Analytic Server V2.0/V2.1 和 V3.0.1 的分析根目录位于不同 MapR 集群上时，已安装的 WEBHDFS 或 NFS 服务将给出指示。需要 WEBHDFS 或 NFS 来复制分析根目录数据，因为无法直接在集群外部访问 MapR 文件系统。

- a. 如果旧的 Analytic Server 2.0/2.1 集群包含 WEBHDFS 服务，请在其中一个新的 Analytic Server 2.1 集群节点上运行以下命令：

```
hadoop distcp webhdfs://<webhdfs_server>:<webhdfs_port>/<old_analytic_root>/analytic-workspace/*
maprfs://<new_analytic_root>/analytic-workspace
```

- b. 如果新的 Analytic Server 3.0.1 集群包含 WEBHDFS 服务，请在其中一个旧的 Analytic Server 2.0/2.1 集群节点上运行以下命令：

```
hadoop distcp maprfs://<old_analytic_root>/analytic-workspace/*
webhdfs://<webhdfs_server>:<webhdfs_port>/<new_analytic_root>/analytic-workspace
```

- c. 当旧集群包含 NFS，且 NFS 也安装在其中一个新的 Analytic Server 3.0.1 集群节点上时，请在其中一个旧的 Analytic Server 2.0/2.1 集群节点上运行以下命令：

```
hadoop distcp file:///<mount_path>/<old_analytic_root>/analytic-workspace/* maprfs://<new_analytic_root>/analytic-workspace
```

- d. 当新集群包含 NFS，且 NFS 也安装在其中一个旧的 Analytic Server 2.0/2.1 集群节点上时，请在其中一个新的 Analytic Server 3.0.1 集群节点上运行以下命令：

```
hadoop distcp maprfs://<old_analytic_root>/analytic-workspace/* file:///<mount_path>/<new_analytic_root>/analytic-workspace
```

查看 MapR 数据迁移站点以了解有关在不同 MapR 集群之间迁移数据的信息。

3. 运行以下命令以更改新分析根目录的所有者和许可权：

```
hadoop fs -chown -R <as_user> <analytic_root>
hadoop fs -chmod -R 755 <>
```

4. 停止 Analytic Server 3.0.1，但请确保元数据数据库仍在运行。

5. 从旧的 Analytic Server 2.0/2.1 集群安装中收集配置设置。

- a. 将 configcollector.zip 归档从新的 Analytic Server 3.0.1 集群安装复制到旧 Analytic Server 2.0/2.1 集群安装上的 <old\_as\_installation\_path>/tools。
- b. 抽取旧的 Analytic Server 2.0/2.1 集群安装上的 configcollector.zip 内容。这样会在旧的 Analytic Server 2.0/2.1 集群安装中创建新的 configcollector 子目录。
- c. 通过从 <old\_as\_installation\_path>/tools/configcollector 运行 configcollector 脚本，在旧的 Analytic Server 2.0/2.1 集群安装中运行配置收集器工具。将产生的压缩 (ZIP) 文件复制到新的 Analytic Server 3.0.1 集群安装。

6. 通过运行 migrationtool 脚本并传递压缩文件（先前由配置收集器创建）的路径作为参数，以在新的 Analytic Server 3.0.1 集群上运行迁移工具。例如：

```
migrationtool.sh /opt/ibm/spss/analyticserver/3.0/ASConfiguration_2.1.0.0.xxx.zip
```

7. 启动 Analytic Server 3.0.1。

---

## MapR 故障诊断

本部分描述一些常见的 MapR 安装和配置问题以及如何解决这些问题。

### hdfsUpdate.sh 脚本的问题

hdfsUpdate.sh 脚本必须只能运行一次，因为该脚本会在上载新文件之前移除分析根目录中的所有文件。如果该脚本运行多次，必须重新上载数据库回送、Hive、HBase 和 Spark 的依赖关系。运行以下命令以重新上载所需的依赖关系：

```
cd <as_installation_path>/jdbc
```

```
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
```

```
cd /opt/mapr/hive/hive-1.2/lib
```

```
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
cd /opt/mapr/hive/hive-1.2/hcatalog/share/hcatalog
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath

cd /opt/mapr/hbase/hbase-0.98.12/lib
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath

cd <as_installation_path>/ae_wlpserver/usr/servers/aeserver/modules/spark/
hadoop fs -put *.jar <as_analytic_root>/cluster1/classpath
```

---

## 第 5 章 配置 IBM SPSS Modeler 以用于 IBM SPSS Analytic Server

为了使 SPSS Modeler 能够与 Analytic Server 配套使用，您需要更新 SPSS Modeler 服务器安装。

1. 配置 SPSS Modeler 服务器以将其与 Analytic Server 安装相关联。
  - a. 编辑主服务器安装目录的 config 子目录中的 options.cfg 文件，并添加或编辑以下行：

```
as_ssl_enabled, {Y|N}
as_host, "{AS_SERVER}"
as_port, PORT
as_context_root, "{CONTEXT-ROOT}"
as_tenant, "{TENANT}"
as_prompt_for_password, {Y|N}
as_kerberos_auth_mode, {Y|N}
as_kerberos_krb5_conf, {CONF-PATH}
as_kerberos_krb5_spn, {AS-SPN}
```

### **as\_ssl\_enabled**

如果在 Analytic Server 上配置安全通信，那么指定 Y；否则，指定 N。

### **as\_host**

托管 Analytic Server 的服务器的 IP 地址。

### **as\_port**

Analytic Server 侦听的端口（缺省情况下为 8080）。

### **as\_context\_root**

Analytic Server 上下文根（缺省情况下为 analyticserver）。

### **as\_tenant**

SPSS Modeler 服务器安装所属的租户（缺省租户为 ibm）。

### **as\_prompt\_for\_password**

如果使用与 Analytic Server 上使用的相同用户和密码认证系统配置 SPSS Modeler 服务器，那么指定 N；例如，使用 Kerberos 认证时。否则，请指定 Y。

以批处理方式运行 SPSS Modeler 时，将 `-analytic_server_username {ASusername}`  
`-analytic_server_password {ASpassword}` 作为参数添加到 `clem` 命令。

### **as\_kerberos\_auth\_mode**

指定 Y 可以从 SPSS Modeler 启用 Kerberos SSO。

### **as\_kerberos\_krb5\_conf**

指定 Analytic Server 应该使用的 Kerberos 配置文件的路径；例如，`\etc\krb5.conf`。

### **as\_kerberos\_krb5\_spn**

指定 Analytic Server Kerberos SPN；例如，`HTTP/ashost.mydomain.com@MYDOMAIN.COM`。

- b. 重新启动 SPSS Modeler 服务器服务。

要连接到已启用 SSL/TLS 的 Analytic Server 安装，需要一些进一步的步骤来配置 SPSS Modeler 服务器和客户机安装。

- a. 浏览至 `http{s}://{HOST}:{PORT}/{CONTEXT-ROOT}/admin/{TENANT}` 并登录到 Analytic Server 控制台。

- b. 从浏览器下载证书文件并将其保存到您的文件系统。
- c. 将证书文件添加到 SPSS Modeler 服务器和 SPSS Modeler 客户机安装的 JRE。可以在 SPSS Modeler 安装路径的 `/jre/lib/security/cacerts` 子目录中找到要更新的位置。
  - 1) 请确保 `cacerts` 文件不是只读的。
  - 2) 使用 Modeler 随附的密钥工具程序 - 可以在 SPSS Modeler 安装路径的 `/jre/bin/keytool` 子目录中找到。

运行以下命令

```
keytool -import -alias <as-alias> -file <cert-file> -keystore "<cacerts-file>"
```

请注意，`<as-alias>` 是 `cacerts` 文件的别名。您可以使用任何喜欢的名称，只要名称对于 `cacerts` 文件唯一。

因此，示例命令将如下所示。

```
keytool -import -alias MySSLCertAlias -file C:\Download\as.cer  
-keystore "c:\Program Files\IBM\SPSS\Modeler\{ModelerVersion}\jre\lib\security\cacerts"
```

- d. 重新启动 SPSS Modeler 服务器和 SPSS Modeler 客户机。
2. [可选] 如果您打算对含有 Analytic Server 数据源的流中的 R 模型进行评分，那么请安装 IBM SPSS Modeler - Essentials for R。IBM SPSS Modeler - Essentials for R 可供下载 (<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>)。

---

## 第 6 章 故障诊断

本部分描述一些常见的安装和配置问题以及如何解决这些问题。

### 常见问题

安装成功但带有警告，而用户无法创建数据源，错误为“无法完成请求。原因：权限被拒绝”

将 `distrib.fs.root` 参数设置为 Analytic Server 用户（缺省情况下，`as_user`）无访问权的目录将导致错误。确定授权 Analytic Server 用户读写以及执行 `distrib.fs.root` 目录。

**Analytic Server 性能逐渐下降。**

Analytic Server 性能不符合预期时，请除去所有 `*.war` 文件，这些文件位于 Knox 服务的部署路径：`<KnoxServicePath>/data/deployments`。例如：`/usr/iop/4.1.0.0/knox/data/deployments`。

### 特定 Hadoop 发行版的问题

在 Hortonworks 2.4 上，禁用 Analytic Server 服务的刷新操作

要手动刷新 Hortonworks 2.4 上的 Analytic Server 库，请按以下步骤操作。

1. 以 Analytic Server 用户（缺省情况下，`as_user`）身份登录到运行 Analytic Metastore 的主机。

注：您可以从 Ambari 控制台找到此主机名。

2. 在 `{AS_ROOT}/bin` 目录中运行 `refresh` 脚本；例如：

```
cd /opt/ibm/spss/analyticsserver/3.0.1/bin
./refresh
```

3. 在 Ambari 控制台中重新启动 Analytic Server 服务。

从外部站点下载的包在 Cloudera Manager 中进行散列检查时失败。

散列验证错误将显示在包列表中。完成下载过程，然后通过 `cloudera-scm-server` 服务重新启动 Cloudera，便可以解决此问题。重新启动服务后，将不会发生此错误。

### 元数据存储库的问题

在运行 `add_mysql_user` 脚本时，操作 `CREATE USER` 失败

在运行 `add_mysql_user` 脚本之前，您首先需要从 MySQL 数据库中手动移除尝试添加的用户。您可以通过 MySQL Workbench UI 或 MySQL 命令移除这些用户。例如：

```
mysql -u root -e "DROP USER '$AEDB_USERNAME_VALUE'@'localhost';"
mysql -u root -e "DROP USER '$AEDB_USERNAME_VALUE'@'$METASTORE_HOST';"
mysql -u root -e "DROP USER '$AEDB_USERNAME_VALUE'@'%';"
```

在上述命令中，将 `$AEDB_USERNAME_VALUE` 替换为您移除的用户名，将 `$METASTORE_HOST` 替换为您在其中安装数据库的主机名。

### 在 Spark 进程中运行的 IBM SPSS Modeler 流的问题

在 Spark 进程中强制运行 SPSS Modeler 流时失败。运行失败的 SPSS Modeler 流与 Analytic Server 源节点（HDFS 文件）一同构建，且已将此源节点链接到 Sort 节点，并设置为导出到其他 Analytic Server 数据源。运行该流后，资源管理器用户界面将指示新应用程序正在运行，但该流不会结束且保持 Running 状态。在 Analytic Server 日志、YARN 日志或 Spark 日志中没有指示流失败原因的消息。

在 Analytic Server 配置中，将 spark.executor.memory 设置添加到定制的 analytics.cfg 文件中，便可以解决此问题。将内存值设置为 4GB，这样可以使先前失败的 SPSS Modeler 流在 2 分钟内完成（在单节点集群环境中）。

## 高可用性集群

由于依赖关系更改，无法将 **Analytic Server** 添加到更多主机

使用第 14 页的『更新客户机依赖关系』中的指示信息运行 update\_clientdeps 脚本。

**java.net.SocketTimeoutException: 读超时**

更改 Liberty ND 超时环境变量，如下所示：

```
export LIBERTYND_READ_TIMEOUT=<milliseconds>
```

其中 <milliseconds> 是用于 JMX 读超时的秒数。

**java.io.IOException: CWWKX7202E: 命令 ./server start 的超时值 60（秒）已到期**

将以下内容添加到控制器服务器 server.xml 中

```
<!-- Increase start and stop server timeout to accommodate slow hardware -->
<serverCommands startServerTimeout="120" stopServerTimeout="120"/>
```

**java.lang.OutOfMemoryError: Java 堆空间**

将以下行添加到 HA 集群的每个成员中的 jvm.options 内。

```
-Xms512M
-Xmx2048M
```

"分析集群服务与 Zookeeper 意外失去联系，将终止此 JVM 以保证集群完整性。"

导致此问题的一个原因是要写入到 Zookeeper 的数据量是否过大。如果，在 Zookeeper 日志中存在类似以下异常：

```
java.io.IOException: Unreasonable length = 2054758
```

或者在 Analytic Server 日志中存在类似以下消息：

```
Caused by: java.io.UTFDataFormatException: encoded string too long: 2054758 bytes
at java.io.DataOutputStream.writeUTF(DataOutputStream.java:375)
```

1. 在 Ambari 控制台中，浏览至 Zookeeper 服务"配置"选项，并将以下行添加到 env-template，然后重新启动 Zookeeper 服务。

```
export JVMFLAGS="-Xmx2048m -Djute.maxbuffer=2097152"
```

2. 在 Ambari 控制台中，导航到 Analytic Server 服务的"配置"选项卡，并在高级分析 JVM 选项中添加以下内容，然后重新启动分析集群服务。

```
-Djute.maxbuffer=2097152
```

要为 jute.maxbuffer 设置指定的数字应大于异常消息中指示的数字。

## Zookeeper 事务数据无法进行管理

将 zoo.cfg 中的参数 **autopurge.purgeInterval** 设置为 1，以实现自动清除 Zookeeper 事务日志。

## 分析集群服务与 Zookeeper 失去联系

查看并修改 zoo.cfg 中的参数 **tickTime**、**initLimit** 和 **syncLimit**。例如：

```
# The number of milliseconds of each tick
tickTime=2000
# The number of ticks that the initial
# synchronization phase can take
```

```
initLimit=30
# The number of ticks that can pass between
# sending a request and getting an acknowledgement
syncLimit=15
```

请参阅 Zookeeper 文档以获取详细信息：<https://zookeeper.apache.org/doc/r3.3.3/zookeeperAdmin.html>

### **Analytic Server 作业未恢复**

存在两种 Analytic Server 作业未恢复的情况。

1. Analytic Server 作业因集群成员失败而失败时，作业通常将在另一集群成员上自动重新启动。如果该作业未恢复，请检查以确保高可用性集群中至少有 4 个集群成员。
2. 停顿某个集群成员时，该服务器上的所有 Analytic Server 作业通常会在另一集群成员上恢复。要确保这些作业能够恢复，请设置 `-Dcom.spss.ae.remoteclient.failover.threshold=100` 并使用远程方式。

### **Analytic Server 服务器偶尔会在服务器关闭时挂起**

手动终止服务器。



---

## 声明

本信息是为在美国国内供应的产品和服务而编写的。可以从 IBM 获取本资料的其他语言版本。但是，您必须拥有该语言的产品或产品版本副本，才能访问对应语言的资料。

IBM 可能在其他国家或地区不提供本文档中讨论的产品、服务或功能特性。有关您所在区域当前可获得的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或默示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务的操作，由用户自行负责。

IBM 可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并不意味着授予用户使用这些专利的任何许可。您可以用书面形式将许可查询寄往：

*IBM Director of Licensing*  
*IBM Corporation*  
*North Castle Drive, MD-NC119*  
*Armonk, NY 10504-1785*  
*US*

有关双字节 (DBCS) 信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：

*Intellectual Property Licensing*  
*Legal and Intellectual Property Law*  
*IBM Japan Ltd.*  
*19-21, Nihonbashi-Hakozakicho, Chuo-ku*  
*Tokyo 103-8510, Japan*

International Business Machines Corporation"按现状"提供本出版物，不附有任何种类的（无论是明示的还是默示的）保证，包括但不限于默示的有关非侵权、适销和适用于某种特定用途的保证。某些管辖区域在某些交易中不允许免除明示或默示的保证。因此本条款可能不适用于您。

本信息可能包含技术方面不够准确的地方或印刷错误。本信息将定期更改；这些更改将编入本信息的新版本中。IBM 可以随时对本出版物中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对任何非 IBM Web 站点的引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：(i) 使其能够在独立创建的程序和其它程序（包括本程序）之间进行信息交换，以及 (ii) 使其能够对已经交换的信息进行相互使用，请与下列地址联系：

*IBM Director of Licensing*  
*IBM Corporation*

North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US

只要遵守适当的条件和条款，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本文档中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际程序许可协议或任何同等协议中的条款提供。

此处引用的性能数据和客户示例仅用于描述目的。实际性能可能因特定配置和操作条件而异。

涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

关于 IBM 未来方向或意向的声明都可随时更改或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

所有 IBM 的价格均是 IBM 当前的建议零售价，可随时更改而不另行通知。经销商的价格可与此不同。

本信息仅用于规划的目的。在所描述的产品上市之前，此处的信息会有更改。

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名称都是虚构的，若实际人员或企业与此相似，纯属巧合。

版权许可证：

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名称都是虚构的，若实际人员或企业与此相似，纯属巧合。

凡这些实例程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明：

© (贵公司的名称) (年)。此部分代码是根据 IBM Corp. 公司的样本程序衍生出来的。

© Copyright IBM Corp. (输入年份)。All rights reserved.

---

## 商标

IBM、IBM 徽标和 [ibm.com](http://ibm.com) 是 International Business Machines Corp.，在全球许多管辖区域的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。最新的 IBM 商标列表可以在 Web 上的 "Copyright and trademark information" 中获取，地址为：[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)。

Adobe、Adobe 徽标、PostScript 以及 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

IT Infrastructure Library 是 Central Computer and Telecommunications Agency 的注册商标，该企业现已成为 Office of Government Commerce 的一部分。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国和其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国和@3B72其他国家或地区的注册商标。

Microsoft、Windows、Windows NT 以及 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

ITIL 是一个注册商标，是 Minister for the Cabinet Office 的共同体注册商标，并且已在 U.S. Patent and Trademark Office 进行注册。

UNIX 是 The Open Group 在美国和/或其他国家或地区的注册商标。

Cell Broadband Engine 是 of Sony Computer Entertainment, Inc. 在美国和/或其他国家或地区的商标并且在当地许可证下使用。

Linear Tape-Open、LTO、LTO 徽标、Ultrium 和 Ultrium 徽标是 HP、IBM Corp 和 Quantum 在美国和其他国家或地区的商标。



Printed in China